Proceedings of the 40th Chinese Control Conference
July 26-28, 2021, Shanghai, China

Prediction Model of Heart Failure Disease Based on GA-ELM

Shaofeng Zhang[1], Weimin Zhou[2]

1. School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444
E-mail: 937257010@qq.com

2. School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444
E-mail: wmzhou@shu.edu.cn

Abstract: In order to predict heart failure disease in advance and keep heart failure patients away from pain, this paper combines genetic algorithms (GA) to optimize Extreme Learning Machine (ELM) and builds a network model that predicts heart failure disease based on some of the characteristic factors that may induce heart failure. The ELM model has the characteristics of fast training speed and strong generalization ability, but its randomly generated input weights and hidden layer bias will cause instability in the prediction results. In this paper, we use GA's principle of survival of the fittest to find the most suitable parameter configuration of the ELM algorithm, and analyzes the simulation results, and compares them with the prediction results of the original ELM model. The results show that the GA-optimized ELM model can better predict the occurrence of heart failure disease and has better predictive performance. The ELM algorithm, which has undergone parameter optimization has made up for its original defects, which may affect clinical practice and become a new assistant for doctors to predict patients with heart failure.

Key Words: Heart Failure, Genetic Algorithm (GA), Extreme Learning Machine (ELM), Simulation Analysis

1 Introduction

Heart failure (HF) is a syndrome in which the pumping function of the heart is impaired due to various reasons, and the cardiac output cannot meet the basic metabolic needs of the body. According to the 2019 China Heart Failure Blue Book Report, it is estimated that there are 6.5 million to 8.75 million patients with heart failure. The human heart is like a transportation hub. It transports oxygen-rich blood to various organs throughout the body through arteries. After all organs are used up, the oxygen-free blood is returned to the heart and then sent to the lungs to replenish fresh oxygen. Heart failure can be divided into left and right heart failures. Once left heart failure occurs, it is as if fresh oxygenated blood in the lungs cannot be sent to the major organs in time, resulting in symptoms related to pulmonary congestion. Right heart failure is because that used and oxygen-free blood cannot be sent to the lungs for replenishment in time, resulting in impaired blood circulation in the limbs.

The prevalence of heart failure increases with age. The prevalence rate of people over 70 years old is as high as 10%, and the mortality rate of patients with severe heart failure can reach 50% within one year [1]. Although the current prevention and treatment programs for the disease have made great progress, the mortality rate in recent years has remained high, which has gradually attracted wide attention from medical workers and patients with heart failure [2]. There are many predisposing factors for heart failure, such as cardiomyopathy, secondary myocardial damage, and cardiac overload [3]. Due to the existence of various predisposing factors, it is not easy to predict the occurrence of heart failure in advance.

For the quantitative prediction of disease progression, clinicians classify heart failure by function, ranging from asymptomatic from ordinary activities to discomfort that may be caused by any physical activity and symptoms at rest. Although this is widely used by peers, this classification still cannot produce a reliable and efficient prediction of heart failure. The importance of the heart organ to mankind is self-evident. If the heart failure event can be predicted in advance based on certain conditions, it will be good news to society. The traditional single-factor prediction method has great limitations and cannot be used to predict complex diseases such as heart failure. Therefore, the feasibility classification method of data based on multiple factors will be the research direction of scholars. Machine learning is currently one of the popular methods used to predict data accuracy, such as support vector machine, BP neural networks, random forest [4], etc. However, the above methods all have their own shortcomings. For example, support vector machine is hard to implement for large-scale training samples, and requires high parameter and kernel function selection; random forest may not produce good classification results when processing low-dimensional data. All in all, it is the key to find a new method that can quickly reach the global optimum, good classification effect and excellent generalization effect to solve the problem of this article.

Extreme learning machine (ELM) has been widely used in various fields since it was proposed. It has faster training speed, strong generalization ability and good accuracy, but its advantage is that it is built in its network through random generation Input weights and hidden layer biases, which also produce the defect that the prediction performance depends on "randomness" and unstable prediction results. Genetic algorithm (GA) has strong robust performance and better global search capabilities. This article will use genetic algorithm (GA) to optimize the input layer weight and hidden layer bias of the ELM network, and establish a new model of heart failure prediction model. The model test data is based on the heart failure clinical record data set in the UCI database, and the validity and feasibility of the model have been verified through multiple comparison experiments.

2 GA-ELM Prediction Model

2.1 ELM Method

Extreme Learning Machine (ELM) is used to train Single-Hidden Layer Feedforward Neural Networks (SLFN). Unlike traditional SLFN training algorithms, the Extreme Learning Machine randomly selects input layer weights and hidden layers Bias. The weight of the output layer is calculated and analyzed based on the Moore-Penrose (MP) generalized inverse matrix theory by minimizing the loss function composed of the training error term and the regular term of the output layer weight norm. Theoretical research shows that even if hidden layer nodes are randomly generated, ELM still maintains the general approximation capability of SLFN. In the past ten years, the theory and application of ELM have been extensively studied. From the perspective of learning efficiency, extreme learning machines have the advantages of fewer training parameters, fast learning speed, and strong generalization ability.

The core idea of the extreme learning machine (ELM) algorithm is to use the traditional single hidden layer feedforward neural network (SLFNs) model idea to transform the training process into a least square solution to solve the problem [5]. The extreme learning machine randomly selects the weight of the input layer and the bias of the hidden layer, and then solves the weight of the output layer, and then calculates and solves it according to the Moore-Penrose generalized inverse matrix calculation theory. ELM algorithm has been widely used and researched after it was proposed. Because of its unique random feature mapping and linear parameter solving process, it has higher learning efficiency and stronger generalization ability. Compared with today's deep learning convolutional neural network, it has fewer training parameters and faster training speed [6] - [9].

The principle of ELM is as follows:

First, we assume that the training set is $\{x_i, t_i | x_i \in R^D, t_i \in R^m, i = 1,2,\cdots,N\}$, where x_i represents the i-th training data, t_i represents the label corresponding to the i-th data, and the overall set is the training set data, which is the input x of the neural network. The network from the input to the hidden layer and then to the output layer are all connected in a fully connected manner. We denote the output of the hidden layer as $H(x)$, as shown in formula (1).

$$H(x) = [h_1(x), \cdots, h_L(x)] \tag{1}$$

Each node $h_i(x)$ in the hidden layer output $H(x)$ is the output of the i-th hidden layer node, so $h_i(x)$ can be expressed by formula (2):

$$h_i(x) = g(w_i, b_i, x) = g(w_i x + b_i), \quad w_i \in R^D, b_i \in R \tag{2}$$

Among them, $g(w_i, b_i, x)$ (w_i and b_i are hidden layer node parameters) is the activation function, which is a non-linear piecewise continuous function that can be used in ELM general approximation ability theorem, commonly used such as Sigmoid function, $tanh$ function, sin function, etc. In this article, we use the Sigmoid function, as shown in formula (3):

$$g(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

After the hidden layer is the output layer, that is, in a broad sense, the output of the ELM algorithm can be expressed as formula (4):

$$f_L(x) = \sum_{i=1}^{L} \beta_i h_i(x) = H(x)\beta \tag{4}$$

Where, $\beta = [\beta_1, \cdots, \beta_L]^T$ is the output weight between the hidden layer (L nodes) and the output layer (m nodes, $m \geq 1$).

The weight w and bias b in the ELM algorithm are generated randomly, which is why it is more efficient than the traditional BP neural network. When the weight and bias are determined, the hidden layer output H can be calculated according to formula (1) and formula (2). In order to obtain the output layer weight β and make the training error as small as possible, we can use the output $H\beta$ in formula (4) and the sample label T to minimize the squared difference as the objective function, and the solution with the smallest objective function is the optimal solution. The objective function can be expressed by formula (5):

$$\min \left|\left|H\beta - T\right|\right|^2, \beta \in R^{L \times m} \tag{5}$$

Where, T is the label target matrix of the training data, and H is the output matrix of the hidden layer. Through line substitution and matrix theory, we can derive the optimal solution as formula (6):

$$\beta^* = H^+ T \tag{6}$$

Where, H^+ is the Moore-Penrose generalized inverse matrix of the hidden layer output matrix H.

2.2 GA-ELM Method

The traditional extreme learning machine has the characteristics of high efficiency and strong generalization ability, but it still has some defects. Because it randomly generates weights and hidden layer node biases when inputting, the prediction accuracy during training will fluctuate, so generating weights and biases that can make the prediction effect is a feasible optimization direction. Genetic Algorithm (GA) is a random search optimization algorithm derived from the laws of natural genetic and biological evolution [10]. It satisfies the idea of survival of the fittest and elimination of the unfit, and covers processes such as gene selection, crossover, and mutation [11]. It has strong robustness and good global search ability. In this article, GA algorithm can be used to optimize the weight matrix and paranoia matrix in the ELM model to achieve the effect of improving the prediction accuracy. Figure 1 shows the overall flow of the GA-ELM algorithm, which includes the initialization of parameters, the optimization process of the genetic algorithm, and the final network prediction.

Number section and subsection headings consecutively in Arabic numbers and type them in bold. Avoid using too many capital letters. If any further subdivision of a subsection is needed the titles should be 10 point and flushed left.

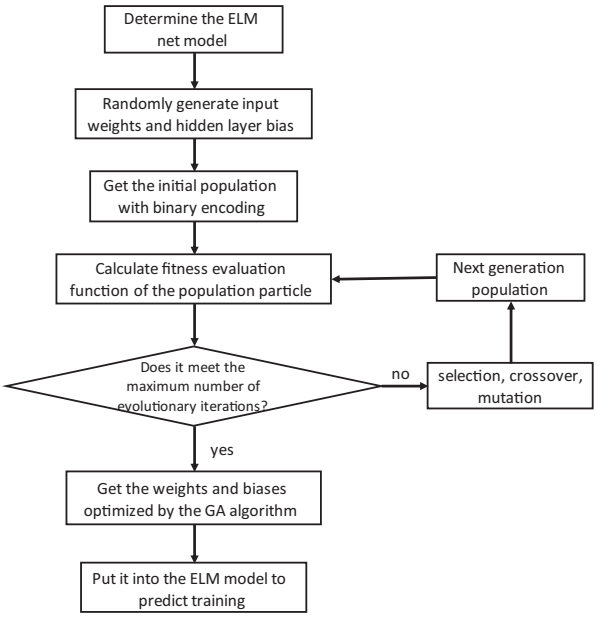

Fig. 1: Algorithm flow chart of GA-ELM

The algorithm idea of GA-ELM is as follows:

Step 1: Population Initialization.

First, let us determine the overall structure of the ELM model, namely the parameter settings of the input layer, hidden layer and output layer. Next, set the maximum number of evolutionary iterations epoch of the GA algorithm, and perform binary encoding on the input weights and hidden layer biases in the randomly generated ELM network, which is the initial population. Each particle in the population is composed of input weight and hidden layer bias, and the length is $D = L \cdot (i + 1)$, where L is the number of hidden layer nodes and i is the number of input layer neurons, namely the dimensions of each vector of the input data. Each particle φ^k in the population is expressed by formula (7):

$$\varphi^k = [\alpha_{11}^k, \alpha_{12}^k, \cdots, \alpha_{1L}^k, \alpha_{21}^k, \alpha_{22}^k, \cdots, \alpha_{2L}^k, \cdots, \alpha_{n1}^k, \alpha_{n2}^k, \cdots, \alpha_{nL}^k, \beta_1^k, \beta_2^k, \cdots, \beta_L^k,] \quad (7)$$

Where, φ^k is the k-th ($1 \leqslant k \leqslant N$) particle in the initial population, and N is the total number of particles; α_{nL}^k and β_L^k are numbers randomly generated in [-1,1].

Step 2: Fitness Evaluation Function.

The ELM model algorithm is used to calculate any particle in each generation of the population to obtain the output weight matrix, and the predicted result and the actual result are calculated for the root mean square error to obtain the fitness evaluation function of the GA algorithm, as shown in formula (8):

$$FitFunc = \sqrt{\frac{\sum_{k=1}^{N}(y_k - y)^2}{N}} \quad (8)$$

Among them, N is the total number of sample particles, k is a single particle, y_k is the predicted output result of the k-th particle, and y is the actual result of the sample. Obviously, the optimization direction of the model is to make the value of the evaluation function as small as possible.

Step 3: Evolution and Iteration.

According to the fitness evaluation function, the smallest particle is selected as the optimal solution. Then initialize the population to perform selection, crossover, and mutation operations based on a given probability, and then optimize the individual to generate a new offspring population until the maximum number of iterations is met. At the same time, we also get the optimized ELM input weight and hidden layer bias.

Step 4: Model Training.

We binary code the weight matrix and hidden layer bias after reaching the maximum number of evolutionary iterations, and input them into the ELM model for training, and finally get the output weight matrix of the improved ELM algorithm.

3 Application of GA-ELM in the Prediction of Heart Failure

3.1 Acquisition of Sample Data

The data set in this article is from the case information of 299 heart failure patients collected in the UCI database in 2015. According to the analysis paper of Davide Chicco and Giuseppe Jurman [12] based on this data set, the research pointed out that the two indexes of serum creatinine and ejection fraction are the most relevant features that determine the cause of heart failure. However, additional studies have shown that patient age, creatinine phosphokinase, serum sodium, and follow-up time are also important factors that may cause heart failure. Therefore, this article will take the six characteristics of the data set (serum creatinine, ejection fraction, patient age, creatinine phosphokinase, serum sodium, and follow-up time) as input parameters. Table 1 is a display of part of the sample data set of six characteristic factors that are most likely to induce heart failure.

Table 1: Display of Partial Sample Data

Feature	1	2	3	4
Serum Creatinine (mg/dL)	1.9	1.1	1.3	1.9
Ejection Fraction (%)	20	38	20	20
Age	75	55	65	50
Creatinine Phosphokinase (mcg / L)	582	7864	146	111
Serum Sodium (mEq/L)	130	136	129	137
Follow-up Time (day)	4	6	7	7

3.2 Parameter Selection

Before training, we divide the data set into a training set and a test set at a ratio of 5:1, and perform de-averaging and variance normalization on the training set data. This process is carried out for each feature dimension, not for a sample.

In view of the parameter settings in the ELM model, we only need to pay attention to the number of hidden layer neurons n and the selection of the activation function, so I studied these two factors in the experiment. Figure 2 shows the effect of the number of hidden layer nodes on the classification performance of the GA-ELM model under the three activation functions (sigmoid, tanh, sin). It can be clearly seen in the figure that when the number of hidden layer nodes is 40, using the sigmoid activation function can obtain an accuracy of 0.786 on the training set. Although the classification effect of using the tanh activation function when the number of hidden layer nodes is 170 is almost the same as the former, the increase in the number of nodes will also increase the amount of training, which in turn reduces the classification efficiency and the training speed. Therefore, this paper selects the activation function as sigmoid, and the number of hidden layer nodes is set to 40. In order to facilitate the comparison of the performance difference between the traditional ELM algorithm and the improved GA-ELM algorithm, this article will select the same activation function and the number of hidden layer nodes.

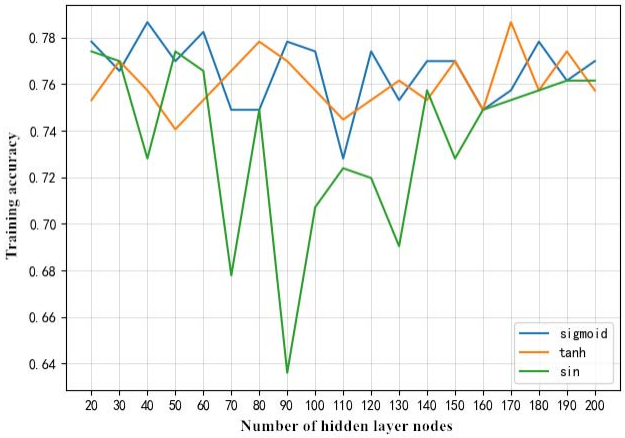

Fig. 2: Influence of the number of hidden layer nodes on classification performance under three activation functions

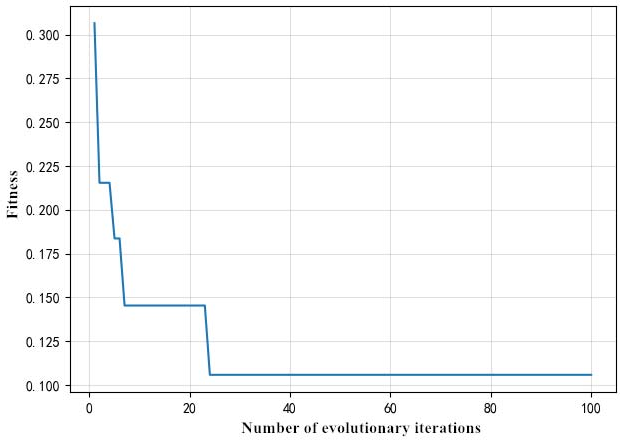

Fig. 3: The fitness change of the GA-ELM during 100 iterations

For the parameters of the genetic algorithm, this article sets the population size to 300, the maximum number of evolutionary iterations to 100, the crossover probability to 0.9, and the mutation probability to 0.1. Then we input the samples into the GA-ELM model for simulation training. From the fitness evolution curve of the model in Figure 3 after 100 iterations, it can be known that when the number of evolution iterations reaches 24, the fitness of the model reaches the lowest value of 0.1058. At this point, we can substitute the input weights and hidden layer biases after 24 iterations into the ELM model for classification performance testing.

3.3 Prediction Result Analysis

Table 2 shows the prediction results of the classification experiments of the traditional ELM model and the improved GA-ELM model on the heart failure patient data set. In the table, the accuracy of traditional ELM on the training set is 76.2%, but testing accuracy is reduced by one percentage point. Although the GA-ELM algorithm combined with genetic algorithm and optimized parameters only improves the accuracy rate of 2.4% on the training set, but its performance on the testing set is 8.3% higher than the traditional ELM, and the accuracy rate reaches 83.5 %. In addition, GA-ELM has greatly improved the recall rate and accuracy rate. Due to the combination of genetic algorithm, time loss will inevitably occur in the process of optimizing parameters, so the last column in the table also shows that GA-ELM consumes more time than ELM in terms of training time.

Table 2: Prediction Results of ELM and GA-ELM

Model	ELM	GA-LM
Testing Accuracy	0.762	0.786
Training Accuracy	0.752	0.835
Recall Rate	0.752	0.835
Precision	0.741	0.846
Time Consuming(s)	0.014	0.522

The prediction confusion matrix of the traditional ELM algorithm and the GA-ELM model is shown in Table 3. We can see that although the traditional ELM model predicts that the probability of final death from heart failure reaches 90%, its predictive probability of survival is only 63%. The prediction results of the GA-ELM model are more in line with people's expectations, the probability of predicting death has dropped, but the predicted probability of survival has reached 79%. This shows that the new model optimized by GA can be better utilized.

Table 3: Prediction Confusion Matrix of ELM and GA-ELM

Prediction Result	ELM		GA-ELM	
	Death	Survival	Death	Survival
Death	0.90	0.10	0.88	0.12
Survival	0.37	0.63	0.21	0.79

The selection of input feature parameters in this paper is based on the research papers of Davide Chicco and Giuseppe

Jurman [12]: In addition to the two most probable features (serum creatinine and ejection fraction), the following are the patient's age, Creatinine phosphokinase, serum sodium. At the end of the paper, they also proved that the factor of follow-up time can be also used to predict heart failure. Therefore, this article uses the six characteristic factors mentioned above to make model predictions.

Table 4: Comparison of GA-ELM with other algorithms

Algorithms	Accuracy(%)
SVM	75.0
Decision Tree	79.6
Random Forest	81.2
k-Nearest Neighbor	81.3
GA-ELM	83.5

This paper is based on SVM support vector machine, decision tree, random forest, K nearest neighbor algorithm and the GA-ELM algorithm proposed in this paper to make predictions. As shown in Table 4, we can see that the accuracy rate increases gradually from top to bottom, and finally get the best prediction effect (83.5%). The GA-ELM model in this paper has perfected its shortcomings because it combines the advantages of ELM, so it has a faster training speed and better generalization ability than other models, and the prediction results also confirm its usability.

4 Conclusion

The research and experimental results in the second section show that the model proposed in this paper using genetic algorithm (GA) to optimize the extreme learning machine (ELM) has the characteristics of good prediction effect and fast training speed.

This paper uses the 'survival of the fittest' rule of genetic algorithm to optimize the input weights and hidden layer parameters of the ELM network. The prediction performance is compared in the simulation experiment. Experimental data shows that the prediction accuracy rate of traditional ELM can only reach 75.2%, while the prediction accuracy rate of GA-ELM model can be increased to 83.5%. It can be seen that the parameters optimized by GA can make the ELM model more stable and reliable, and then produce better prediction results.

The GA-ELM model provided in this article can achieve better prediction results of heart failure disease through limited data accumulation, which provides doctors and potential heart failure patients with the possibility of timely diagnosis of the disease. Of course, GA-ELM is not perfect. We still need much more experimental data to prove its feasibility. The follow-up work will be to find more sample data sets to train and optimize the model. If necessary, we will try more advanced models to predict heart failure disease.

References

[1] Li Xiaoying. Survival prediction model for patients with heart failure[J]. Chinese Journal of Cardiovascular Diseases, 2009(06):569-572.

[2] Shen Wenjia, Yang Qiaofang. Study on the current status and influencing factors of the pre-established medical care plan for patients with chronic heart failure[J]. Practical Journal of Cardiovascular and Cerebral Pneumal Disease, 2020, 28(10): 102-106.

[3] Koushik Reddy,Asma Khaliq,Robert J Henning.Recent advances in the diagnosis and treatment of acute myocardial infarction[J].World Journal of Cardiology,2015,7(05):243-276.

[4] Liu J,Li M,Liu R R,Zhu Y,Chen G Q,Li X B,Geng C,Wang J J,Gao Q X,Heng H Y. [Establishment of a CT image radiomics-based prediction model for the differential diagnosis of silicosis and tuberculosis nodules].[J]. Zhonghua lao dong wei sheng zhi ye bing za zhi = Zhonghua laodong weisheng zhiyebing zazhi = Chinese journal of industrial hygiene and occupational diseases,2019,37(9).

[5] Zeng Lulu. Application research of improved PSO optimized ELM algorithm in medical image segmentation [D]. Guangxi Normal University, 2016.

[6] Shuo Meng,Jianshe Kang,Kuo Chi,Xupeng Die. Gearbox fault diagnosis through quantum particle swarm optimization algorithm and kernel extreme learning machine[J]. Journal of Vibroengineering,2020,22(6).

[7] Michael Li,Lily D. Li. A Novel Method of Curve Fitting Based on Optimized Extreme Learning Machine[J]. Applied Artificial Intelligence,2020,34(12).

[8] Zhaoxi Wang,Shengyong Chen,Rongwei Guo,Bin Li,Yangbo Feng. Extreme learning machine with feature mapping of kernel function[J]. IET Image Processing,2020,14(11).

[9] Min Yan Chia,Yuk Feng Huang,Chai Hoon Koo. Swarm-based optimization as stochastic training strategy for estimation of reference evapotranspiration using extreme learning machine[J]. Agricultural Water Management,2021,243.

[10] Zheng Maohui, Liu Shaofi. GA optimized ELM neural network for the diagnosis of drainage pipeline defects[J/OL]. Journal of Harbin Institute of Technology: 1-7, 2020-10-30.

[11] Qiqi Zhang,Bo Fan,Kexing Song,Hua Huo,Wenwu Hou. Research on Intelligent Modular Process Recombination based on Improved Genetic Algorithm[A]. SCIence and Engineering Institute (SCIEI),2019:6.

[12] Davide Chicco,Giuseppe Jurman. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone[J]. BMC Medical Informatics and Decision Making,2020,20(1).