Diagnosis of hyperlipidemia in patients based on an artificial neural network with pso algorithm

Asma Naeimi Baghini[1], Minoo Soltanshahi[2], Amir Rajabi[3]

Abstract - One of the most common and most dangerous diseases of blood fats are such as heart disease, diabetes and stroke, heart and brain. It can control the timely diagnosis, treatment and then prevention of complications is become very effective even without using medicine. Heart disease and diabetes file if patients has useful information that can be used to estimate blood fat timely diagnosis. In this paper we introduce a method based on data mining according to the information of patients' medical records to predict and detect blood lipid cardiovascular. And to identify patients with high blood lipids,we use a category based on neural network without feedback and pso algorithm to train the neural network to determine the appropriate value to reduce error the weights of the neural network . Simulation is done in MATLAB environment by using Body Fat data set, it shows the accuracy of 93.22 percent compared to the same methods, which means high accurate, higher detection sensitivity and Democrats .

Index Terms - hyperlipidemia, data mining, neural network algorithm pso, prognosis, cardiovascular disease.

1- Computer Engineering Department, Kerman Branch, Islamic Azad University, Kerman, IRAN.
(As.naeimi@gmail.com)
2- Computer Engineering Department, Kerman Branch, Islamic Azad University, Kerman, IRAN.
3- Computer Engineering Department, Kerman Branch, Payam-Noor University, Kerman, IRAN.

I. INTRODUCTION

Every year many people lose their lives due to heart disease. The origin of heart disease is fatty deposits on blood vessel walls of blood flow to the heart. One of the most common factors such as blood cholesterol and heart diseases, vascular, stroke, diabetes and high blood pressure, kidney failure and so on, That early diagnosis can prevent complications and so control and the treatment are very effective even without using medicine . Early diagnosis helps the blood fat people to reduce blood fat and reduced accumulate in artery walls by apply different methods And reduce the risks of this disease [1]-[4]. The best tool of blood lipid detection which accurately measure blood fat is Blood tests.

Blood tests has a huge barriers, including the unavailability of laboratory, fear of the sample, blood test fees etc. [1]-[3]. Heart disease and diabetes file of patients has useful information that can be used to estimate and predict blood fat. Discovering hidden patterns and information is not possible simply without special tools.

Data mining is a good method in order to discover hidden patterns and medical information diseases in large amounts medical information's. In data mining of information that apparently is not a significant relationship between them different patterns are discovered. Data mining has many applications in medicine. Data mining tools are very diverse, One of the most important tools is data mining of neural networks, Neural network are able to forecast and evaluate data with little error in predicting events and their data by using training sets [1]-[7]. Hence in this article to identify and predict blood lipid neural network, algorithm pso is used to reduce these

errors. One of the challenges in estimating and predicting neural network events is an error between the predicted values and actual values. In fact, a neural network data model was made by blood fat is a prediction model that is based on the training data.

Adapting this model to predict real models predictive models is needed to reduce error this process needs to determine which the correct and optimal weight neural network is. Reduce the amount of error prediction model and real model is an optimization problem. In the proposed method to optimize the accuracy of neural network, particle swarm intelligence is used which is imitate the behavior of birds. In swarm intelligence, each particle itself has little intelligence, but when the particles interact with each other has a great intelligence that could well solve difficult problems [pso]. In this article a combined neural network with particle swarm intelligence algorithm are used to predict the measurement of blood fat, part of it goes on like this: In the second part is the background of the problem , and third and fourth respectively in the relevant literature and proposed method, The fifth and sixth parts are also to evaluate the proposed approach and conclusions.

II. BACKGROUND

Blood fat, or Hyper lipid is term which is used by doctors to describe the high levels of blood fat or fat particles in their blood. Lipid is the scientific term in the case of fat in the body. Body and blood fats have useful benefits such as energy storage, to build cells and useful hormones. Measuring the level of blood plasma lipids can be used to determine and measure blood lipid disease and by given the level of this fat, take the necessary precautions. Important parameters in the diagnosis of blood fats are Cholesterol, LDL, triglyceride, HDL and LDL, that we measure them to used to detect blood fat [1]-[2]. Heart disease is a leading cause of death in the world today, the most important reasons is blockage of arteries supplying blood to the heart or coronary arteries. Scientists in the face of the approximately epidemic of coronary artery disease, have been identified some factors as risk factors. Such as blood fat, the development of coronary artery disease which is based on atherosclerosis, or hardening of the arteries. The disease in addition to heart attacks

responsible for most cases of strokes, many cases of kidney failure and peripheral artery disease that is usually in the hands and feet can lead to disease, gangrene and amputation. In general it can be said Coronary artery disease is the result of several factors such as : Smoking, high blood pressure, high blood cholesterol, diabetes, lack of exercise, obesity, abdominal obesity, unhealthy diet high in fat and high in salt, age, gender, family history, genetics, alcohol consumption, psychosocial factors, stress, menopause and high blood glucose [1].

The best way to check for coronary heart disease is Angiograph, unfortunately angiography is a procedure that is expensive and dangerous and risks such as death, myocardial infarction and stroke hence, the non-dangerous and non-invasive methods are of most interest [2].

III. RELATED WORKS

1. Diagnosis Using Neural Network
In 2012, Mr. Reza Ali Mohammad Portahamtanand and his colleagues used a Multilayer Perceptron Neural Network, with Back propagation algorithm, for evaluation of coronary artery disease, among 150 patients in heart hospital in Mazandaran [28]. They initially found to be an overview of the relevant variable for the statistical analysis conducted quantitative and qualitative data. The mean quantitative variables such as age, creatinine, ejection fraction was significantly different in the two groups of healthy and sick. However, body mass index, cholesterol and triglyceride levels were not significantly different between the two groups. Information on qualitative variables showed that smoking and disease but other variables such as gender, exercise test results and high blood pressure, diabetes, thereby percent, to echocardiography, angiography is associated with significant results [28].

2. Diagnosis using combined data mining models
In 2014, Ms. E. Rahimi Shatranlv et al. in their study, Predict coronary heart disease samples by using the data mining techniques, They were randomly selected 450 cases of coronary disease hospital and study the extraction of relevant variables records in accordance with the methodology provided[1]. In this step, data mining research methodology using predictive

algorithms to predict coronary heart disease, ultimately to improve forecasts, the hybrid model of proposed decision tree and Bayesian Network is suggested. They used records demographic variables Such as age, sex, weight, height, medical backgrounds or like hypertension, hyperlipidemia, diabetes, cigarettes, Laboratory measurements such as total cholesterol, good cholesterol, bad cholesterol and blood sugar, triglyceride as well as the current situation through sickness extraction and collection, to be able to have high-quality network search algorithms used the Bayesian network tan. In compare with two initial models using the search algorithm TAN, We can say that the accuracy of hybrid modeling is more than the initial model. The amount has increased from 0.9 to 0.95. The algorithm combines models has an accuracy of 0.95 is 0.95 [1].

3. compares the performance of decision trees and neural networks in predicting myocardial infarction

In this study, Mr. Safdar and colleagues using neural network algorithms and decision trees to model and extract the rules in order to predict risk of myocardial infarction. we compare the results of similar studies, In Table I [2].

Table I: Comparison of the results of studies in the field of data mining in heart diseases [2]

Authors and year	The algorithm used	Disease	Accuracy	finding	Predictors
Jyoti (2011)	Bayesian network decision trees, artificial neural network	Heart disease	89%	Create rules for relationships between variables	Gender, age, chest pain, high blood pressure, fasting blood sugar, cholesterol levels, smoking, body mass index and ...
Mohammad Pur (2011)	Artificial Neural Network	Coronary heart disease	96%	Correct classification of patients needing cardiac catheterization and pharmacotherapy	Age, body mass index, triglycerides, history of hypertension, history of diabetes, history of heart disease, exercise test results and ...
Biglarian (2004)	Artificial Neural Network, Logistic Regression	Coronary artery bypass graft	99/33%	Better performance, neural networks, in-hospital predictors of mortality after open heart surgery	Age, body mass index, cholesterol, triglycerides, blood pressure, smoking, diabetes, hyperlipidemia, heart disease and ...
Christine (1998)	Logistic regression, classification trees	Myocardial infarction	81%	Decision Trees better performance in preventing myocardial infarction	Age, family history of heart disease, smoking, chest pain, high blood pressure, diabetes, night sweats, vomiting and ...

In a survey conducted by Jyoti for predicting the risk of heart disease by decision tree model was 89 percent accurate. However, that difference can be noted in the study had a greater number of variables [29]. Mr. Mohammad Pour study the

neural network in evaluation of coronary heart disease and precision obtained is equal to 96%. That demonstrate the power of this model is faster diagnosis of patients who require diagnostic and therapeutic treatment, The high sensitivity of the proposed model can be due to the use of variables such as exercise test, and the results echo and also determine the number of neurons in the middle layer neural network is considered less [30].

In a paper titled Application of artificial neural network to determine predictors of in-hospital mortality after open heart surgery and comparison with the logistic regression model, were used an artificial neural network with 18 input neurons, hidden neurons and two neuron output 4 propagation algorithm for evaluating patients who had undergone open heart surgery at the hospital. Its accuracy is 99/33 percent but in the logistic regression model provided 90% accuracy is obtained compared to the neural network becomes clear that the neural network is capable of solving the logistic regression model [30].

Christian in his study to compare the performance of several algorithms decision tree to determine the risk of myocardial infarction, and decision tree model with a sensitivity of 81% to be a suitable model for predicting introduced [31].

4. data mining algorithms to predict heart disease

Data mining of statistical analysis, merges machine learning and database technology to extract hidden patterns and relationships between large data bases. Data mining and modeling work discovered large amounts of data to discover relationships with disciplines that are initially unknown. In order to obtain a clear and useful result for the database. One of the methods that have been used recently is use ant colony in data mining to detect heart disease [32]. Classification of a complex framework of laws Community exploration and standings. WAC is Weight classification techniques. This is a novel idea and dependence for the classification of uses, Different weights to different pages or names assigned in accordance with their ability to predict. The pre-processing and data warehouse heart disease characteristics in the range of 0 to 1 is given a weight that reflects its importance as the model predicted, The adjectives that have greater impact high weight and low weight

are assigned attributes that have less effect, The results showed that WAC is more efficient compared to other methods of classification or modify database and consider two classes instead of five classes, integrated database becomes more and 81.51% predictive accuracy of this method is that the highest accuracy [33]. Studies show that the classification performance of the classifier traditional associative is better. When medical association rules are done on a data set and many of which are irrelevant to medicine so the time required to find them is very high. To solve this problem, solutions such as filtering elements, clustered, the maximum size of elements, filtering the early-Late are proposed. In general association rules are extracted from the input data set be acknowledged without being dependent on a single sample. To solve this problem the author introduces an algorithm that uses rules to reduce the number of search constraints. Find association rules are performed on the training set and at the end of a test series depends admit. Weight classification system can also be used in remote areas such as rural areas. The system is user-friendly and feature updates when new data set will be inserted. Efficiency of this algorithm is 81.51% [33].

5. Decision tree algorithms to predict heart disease

Ms Zamapoor and et al, research was analytical and its database has contains 353 records. The data needed for the study in 2012 were obtained from the records of patients admitted to hospital. In this study for the construction of decision trees and neural network models variables of gender, age and history of smoking, addiction, history of hypertension, history of blood fat, sugar and fat factors, body mass index, blood group is determined as predictor variables and variable or non-disease risk is determined as the target variable. The algorithm used in this study, c5 algorithm has the highest accuracy rate was 93/4 percent. And the most influential factors were age, high blood pressure, high blood fat and smoking. Therefore obtained using the rules for a new person with certain variables, which can be determined how much would be at risk of developing myocardial infarction [3].

IV. THE PROPOSED METHOD

In this section, the proposed method is a developed method with the multi-layered artificial neural network, to detect blood fat with minimal clinical trials and save time and expense recognition, will be presented as well. Advantage of predictable and accurate diagnosis of hyperlipidemia with the least possible error which minimize number of clinical trials. Because of the reduction in costs and the time of diagnosis this software is widely used. Our proposed method is based on the combination of artificial neural network with pso algorithm. Therefore, in the remainder of this section at first neural network algorithm then combines pso and eventually use them in the proposed method is described.

1. Neural Networks

One of the most important tools in data mining, and detect disease patterns neural networks can be cited. Neural networks are suitable tool for modeling complex problem that cannot be solved by other methods or difficult to dissolve. One of the most important ways in the diagnosis and prognosis of the disease is using neural network [34]-[36]. In this case, the neural network using various features which refers to them as input data to offer the space features a regression graph to predict. This regression is a defining characteristic of a two-dimensional diagram in which the horizontal axis represents the feature space and the vertical axis is the output function, neural network with two features represent a curve in three-dimensional space. With increasing the number of features of a problem in n-dimensional space, rather than neural network regression line or two or three-dimensional space, a cloud modeling page that displays it difficult and somewhat is impossible. To create a predictor neural network model and they use the sum of the weight. The weight classes are used for modeling cloud separator page. In Figure 1, at the entrance of each neuron in the form of a circle the weight applied and this weight to the input neurons to multiply and takes the data used in the estimate and forecasts [34]-[39].
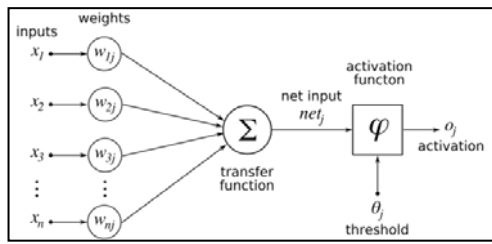
Fig1. Structure of a neural network by applying weights on the inputs and their sums with the related neurons

Suppose that Xi a training record in neural network that has Yi output. This training record is well shown in relation 1.

Relation (1) $X_i = \ll x_{i1}, x_{i2}, \ldots, x_{in} \gg$

In this regard, X_i a record of training data is input. Where n is the number of input column training record. In the relation 2 the column of a training record is shown.

Relation (2) $Y_i = \ll y_{i1}, y_{i2}, \ldots, y_{im} \gg$

In this regard, Y_i a record of training data is output. Where m is the number of columns of output in the training record. If the data input and output, show respectively by the vector χ and γ, neural network is like a diagram 2.

Fig 2. Structure of a neural network and feedback from the errors for improving weight of the layers

In the above figure γ' is prediction and estimation of a neural network that Subtraction of the γ represents an error of neural network. A good neural network has a minimum sum of squared errors; the relation 3 shows these criteria well. Relation (3) is a function error.

Relation (3) $f = e_1^2 + e_2^2 + e_3^2 + \cdots + e_n^2 = \sum_{i=1}^{n} e_i^2$

The aim of this research is to reduce this problem. In fact it's an optimization problem; there is a need to minimize. In this study, using pso algorithm to minimize these criteria, to the model prediction of the actual model may have the least difference [34]-[39].

2. PSO Algorithm

Particle swarm algorithm is an optimization algorithm, mimics the behavior of animal societies in processing knowledge society. This algorithm is derived from two field's .The first artificial life (such as birds, fish) and a second evolutionary computation. The basis of development of Pso algorithm is to consider, possible solutions in an optimization problem-free as birds volume and quality characteristics. This is referred to them as particles[40]. The birds' flight in an n-dimensional space and its path in the search space based on their past experiences and their neighbors change. In such an atmosphere, the assumptions are made and assigned speed elementary particles. The communication channels between the particles are considered. Then the particles are moving in the space and the results are based on a criterion of merit is calculated after each period. Over time, the particles go toward the particles that have higher fitness standards and are in the same communication, accelerated, the main advantage of this method is that the number of particle swarm optimization strategies, the local optimal solution is the flexible approach to the problem. Each particle has a position that defines a multi-dimensional coordinates of the particle in the search space, the particle motion over time will change the position of the particle. Xi (t) shall determine the position of the particle i at time t. Every bit of space to move also requires a speed vi(t) the velocity of the particle i at time t specifies. adding speed to the position of each particle can be considered a new position for the particle. Relation 4 determines the position of the particle[40-41].

Relation (4) $x_i(t + 1) = x_i(t) + vi(t + 1)$

Fitness function is responsible to know whether the position of a particle's position is appropriate search space or not. Particle remembers best position that it had been during his lifetime. The best position to best meet individual experience of a particle by particle named pbest and particles can also be aware of best position visited by the whole group that this position is called gbest. Relation 5 defines the relationship between the bit rates.

Relation(5)
$$v_i(t+1) = v_i(t) + c_1 r_1 \left(pbest_i(t) - x_i(t) \right) + c_2 r_2 \left(gbest_i(t) - x_i(t) \right)$$

Particle velocity vector in the optimization process reflects empirical knowledge and information society particle particles. Each particle in the search space for the two components to consider is[40]-[41]:

1- Cognitive component: The best solution is that a particle acquires alone.

2- The social component: the best solution that is recognized by the entire group.

3. pso use in training neural networks

Optimization variables included training a neural network weights and biases of the network. If you have n layers of a hypothetical network of R and M neurons is input, The matrix Wn weights and biases Bn this layer with relation 6 can be displayed as follows [4]:

Relation(6)

$$wn = \begin{pmatrix} (w_1^n) \\ (w_2^n) \\ . \\ . \\ . \\ (w_M^n) \end{pmatrix}^1 \qquad Bn = \begin{pmatrix} (b_1^n) \\ (b_2^n) \\ . \\ . \\ . \\ (b_M^n) \end{pmatrix}$$

Where $w_m^n = [w_{m1}^n \, w_{m2}^n \, \dots \, w_{mR}^n]^T$ is the vector of weights of m neurons of the input layer to layer my M. Similarly, the per layer weight matrices and bias vectors corresponding parameters are defined. With the following parameter vector all network layers, vector optimization variables to be formed.

In fact, this vector is the position vector mentioned in relation 7, the optimum value argument will be calculated using pso algorithm.

Relation (7) $\qquad X = \begin{pmatrix} x^1 \\ x^2 \\ . \\ x^L \end{pmatrix}$

If the process is to first position N of Xi vector, where N is the number of gang members, randomly generated. Neural network parameters for the vectors of variables run. At this stage vectors pbest, gbest obtained due to the propriety are calculated. And the new position vector n is produced using relation 4 and 5. This process is repeated until the final convergence achieved. The final integration vector achieves optimal position, in a way that minimizes the training error for it. Also, the coefficients c1, c2 = 2 is selected [4]-[41].

4. The proposed approach

At first, the input variables that are described below are trained artificial neural network. The results which are vectors, transforming the incoming particles Pso algorithm and then we minimize the Pso Algorithm Neural network classification error as much as possible. Optimal selection neural network weights and thresholds minimize the average forecast error. To simplify the minimum classification used Pso algorithm computation and neural network through which the optimal weights and thresholds to be determined. And therefore the most precise neural network is used to predict. In general, the proposed method of artificial neural network as a member of the initial population in Pso algorithm displayed as a vector and to assess the merits of neural networks in the prediction of disease and more accurate classification of the average classification error in the data used in the test.

1. Input data sets:

One of the most important topics of data about heart disease and atherosclerosis, is the data set the Cleveland Clinic Foundation, The collection originally had 76 features that further study of fourteen important feature is used. In this data set with 303 samples or records as well. From 14 features, 13 feature are an input dataset and the latest feature, shows the output data set that are explained below.

1) Age
2) Sex

3) types of angina: Angina (chest pain) caused by partial obstruction of a coronary artery heart of the concept is that the heart not getting enough blood. In this data set for field typical angina, atypical angina, painless and asymptomatic angina is considered respectively; with the numbers 1,2,3,4 are shown.

4) Blood pressure at rest

5) Cholesterol or blood plasma

6) Blood glucose levels at breakfast

7) The results of electrocardiography

8) The maximum number ever recorded for a patient's heart rate

9) Exercise-induced angina

10) The indent ST ECG wave

11) ST wave activity in the heart that has three modes steep curve, which is flat and low slope values of 1, 2, 3 are shown.

12) Color imaging of blood vessels seen in the number between 0 and 3.

13) Type of thalassemia: Last input feature dataset that has three values are 3,6,7.

14) The output characteristics of the data set have 5 different classes the possibility of clogging of the arteries due to cholesterol in heart vessels show.

15) This feature is 0, 1,2,3,4 zero values that represent health and number 4 marks the very high risk of Coronary Heart Disease by blood lipids shows.

2. Initialization

We have introduced the important parameters of the initial population equal to 10 and the number of repetitions to 30, and 95% of the input data as training data and the remaining 5% as the test data. The initial value of the evaluation parameters TN, TP, FN, FP is zero and the values of a, b normalization were defined to have the value of -1 and 1. Our proposed method with pso algorithm help to minimize detect errors classification neural network as much as possible and blood fat and data of these patients are used as training data. This data, including weight, height, age, gender, etc. can be cited. One of the challenges in estimating and predicting neural network is an error between the predicted value and the actual value. In fact, in our proposed method is a neural network data model that makes blood fat that is a prediction model was based on the training data.

3. Neural network proposed

The proposed neural network structure based on the future of the data set is made of blood fat and given that blood fat dataset with 13 features input and one output features. The proposed neural network has 13 number input and 1 output that its output is a number between 0 to 4, 0 indicates a healthy number of person and number four very high risk of Coronary Heart Disease by blood fats show. We have defined neural network with two layers, the first layer have 5 neurons and the second layer have 3 neurons. Quality classification and prediction of an artificial neural network classification error by the medium that the objective function is often measured mean square error is calculated. The objective function is shown in equation 8.

$$\text{Relation (8)} \quad e = \frac{1}{n}\sum_{i=1}^{n}(d(i) - y(i))^2$$

Where n is the number of training data, d (i) the actual amount of data i, y (i) the estimated value of i and e mean square error of classification or is anticipated. Minimizing the objective function of an optimization problem is difficult, due to certain pre-assumptions, complexity and nonlinear methods such as gradient function is not resolved. Unlike other evolutionary algorithms for solving optimization method requires no specific default derivatives, such as viability, being continuous, linear or nonlinear objective function is not simply a well solve these issues.

V. THE EVALUATION AND SIMULATION

To implement the proposed algorithm of MATLAB programming environment is used. The software environment for performing numerical calculations and a fourth-generation programming language is due to the enormous possibilities, and having the ability to draw complex graphs of data mining algorithms, a programming environment is efficient.

1. Normalizing the data

In order to get the logical and desirable answer from the model, is necessary before training the network, inputs and outputs are limited to a certain period. The purpose of correction is, reducing network modeling error. This process

is called standardization or normalization. Normalization cased the learning neural network done with better quality and a particular instance can cause significant error in the output of the neural network. In the proposed method used relation 9 to normalize the data values between -1 and 1:

Relation(9) $x_i^{'} = a + \dfrac{x_i - x_{min}}{x_{max} - x_{min}} (b - a)$

Where x_i is i features samples, $x_i^{'}$ features i have normalized samples, x_{min} least I features and x_{max} most characteristic feature and a, b are amount of returns normalized. In the figure below you can see normal output data of entrance. Normalization of 15 first data is shown in Figure3.

Fig 3. The normalized outputs of the first 15 records of the blood lipid dataset

2. Evaluation Criteria
Typically, to measure performance and evaluate multiple data mining algorithms measure of accuracy, sensitivity and detection are used. In order to calculate these criteria require that the concepts of true positive, true negative, false positive and false negative diagnosis of blood fat, and then they calculate as the main criteria for accuracy, sensitivity and detection are used [2]:
1) true positive(TP): the number of blood fat test data is that the proposed approach they correctly diagnosed.
2) true negative(TN): the number of healthy people in the test data that the proposed method is properly diagnosed them healthy.
3) false positive(FP): the number of healthy people in the test data that the proposed approach is wrongly diagnosed them.
4) false negative(FN): The number of blood fat in test data that the proposed approach is wrongly diagnosed them healthy.

Accuracy
This criterion is as the ratio of true positive and true negative samples, all samples are defined as the relation 10 shown.

Relation(10)

$$Accuracy = \dfrac{TP + TN}{TP + FP + TN + FN}$$

Sensitivity
Sensitivity to the relationship 11 is defined as the number 1which represents the desire of the proposed algorithm effectiveness.

Relation(11) $Sensitivity = \dfrac{TP}{TP + FN}$

Specificity
Recognition in the form of Equation 12 is defined as the desire to increase the number '1' indicates that the proposed algorithm is effective.

Relation(12) $Specificity = \dfrac{TN}{TN + FP}$

The sensitivity can be proposed as algorithm that is able to detect blood lipid disease and consider feature the ability to identify healthy individuals [2].

3. The results of the simulation
In this section, for better evaluation, at each stage of the simulation, a constant parameter and the effect of changes of other parameters be checked.

3.1 Effect of initial population size of the forecast error
Here are two important parameters such as the number of repetitions and the initial population is shown. Each of the figures for 20 repeats and population size 20, 30, 40 and 50 are shown. As output figures 4, 5, 6 and 7 shows, by increasing the number of initial population, are more likely to eventually minimize the error rate increases.
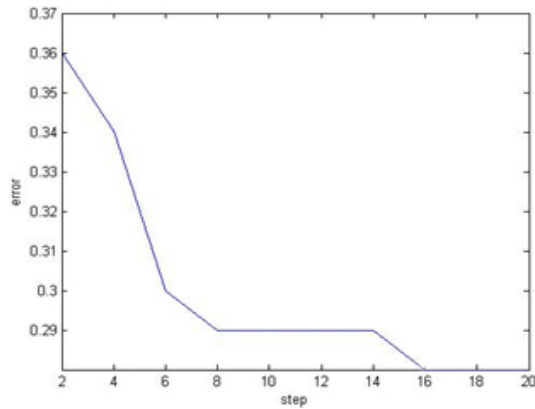
Fig 4. The average error for the initial population is 20 and the number of repetitions 20
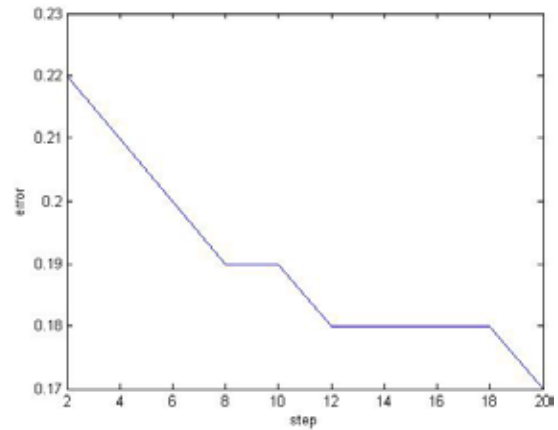

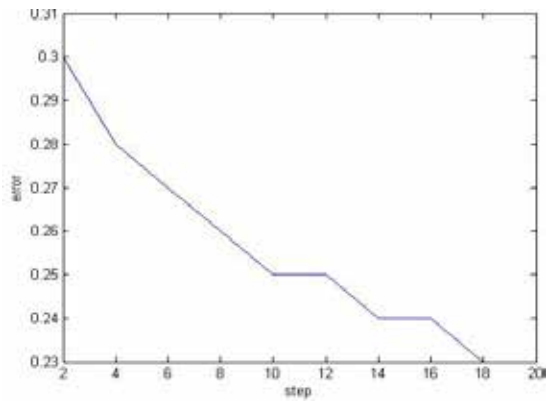**Fig 7. The average error for the initial population is 50 and the number of repetitions 20**

Fig 5. the average error for the initial population is 30 and the number of repetitions is 20

In these figures, the accuracy is calculated respectively as 0.936, 0.946, 0.955 and 0.965. On the whole, it can be concluded that increasing the number of the initial population leads to the increase in accuracy of the classification of the blood fat disease in the proposed method.

3.2 Effect of repeated impact on prediction error

In each Figure of 8 and 9, the classification error for the initial population of 30 patients with hyperlipidemia and repetition of 20 and 30 variables is shown.
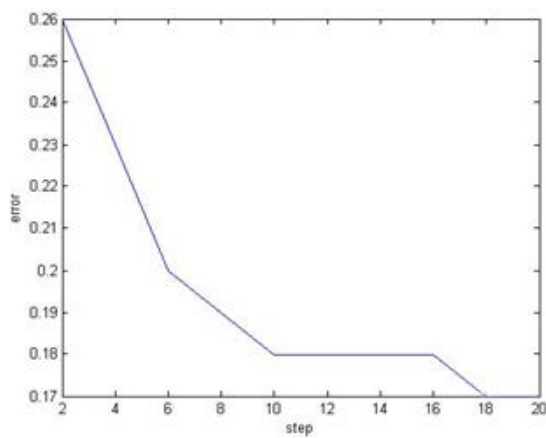

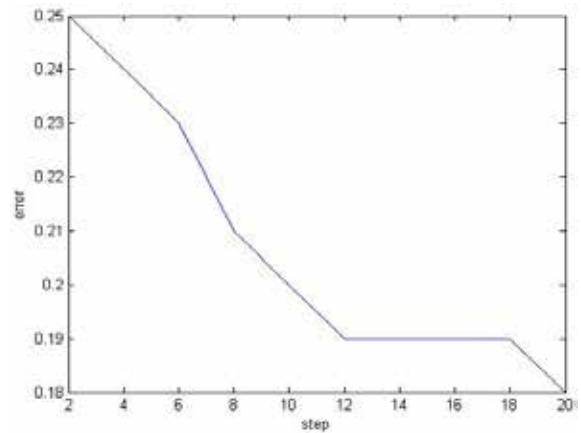**Fig 6. The average error for the initial population is 40 and the number of repetitions 20**

Fig 8. The average error for the initial population is 30 and the number of repetitions 20
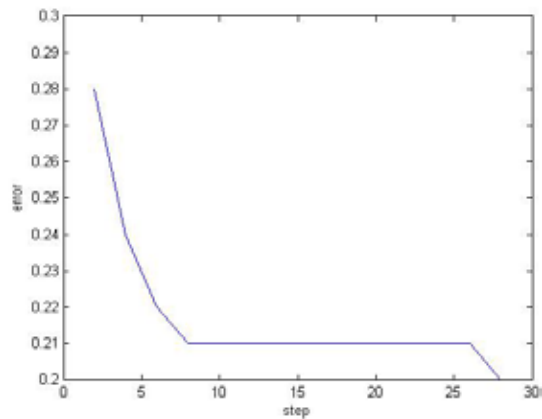
Fig 9. The average error for the initial population is 30 and the number of repetitions 30

In these figures, the average accuracy is achieved 0.962, 0.957 respectively. As output of the charts and two forms of output show, increase in repetition is more likely to eventually increase minimization of error rate. In general it can be concluded that increasing the frequency of the proposed algorithm increases accurately classify lipid patients in the proposed method.

3.3 comparison of the results

Mr. Reza safdari and colleagues [2] released article comparing the performance of decision trees and neural network in the prediction of myocardial infarction, with the closest match to the subject of this study. We discussed this step after the creation of the proposed model to evaluate it. Verify the data model into two categories: education (80%) and exam (20%) were divided. The model can be built by the educational sector. Data on the section test, evaluates the model. Also, to evaluate models indicators of sensitivity, specificity and accuracy can be used. The results obtained for the proposed method are tabulated and shown in Table II and Figure 10.

Table II. The comparison of the proposed method with the neural network and the decision tree

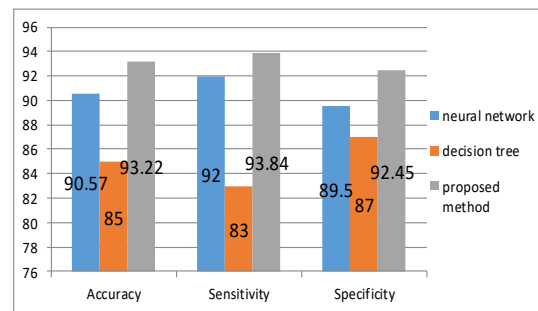|	proposed method	decision tree	neural network
Accuracy	93/22%	85%	90/57%
Sensitivity	93/84%	83%	92%
Specificity	92/45%	87%	89/5%

Fig 10. Comparison of proposed with neural network and decision tree method

Also, in order to have better assess, in the neural network combined approach pso, we also simulated the neural network alone. It can be said that compared two models predict that the hybrid model has higher accuracy and sensitivity. Accuracy is 93/22 percent in the hybrid model and sensitivity is 93/84 percent, the criteria in neural network modeling 90/57 percent and 92 percent respectively. In addition to the features in the hybrid model is better and higher.

VI. CONCLUSION

In the proposed method, the efficiency is the most dependency to the number of initial population and the number of iteration. This means that by increasing the initial population size and the number of repetitions, the classification sick people error is reduced and the results showed that the accuracy, sensitivity and specificity of the proposed method, respectively, with values of 99/22%, 93/84%, 92/42% respectively. Compared to the neural network and decision tree shows better values and neural network combined with pso algorithm increases the accuracy of detection neural network. This result is very important because of the complications and potential damage angiography for patients who do not need it can be avoided. On the other hand, it can diagnosis the patients who really need treatment in the quickest time with the greatest accuracy. This knowledge can be found in health centers to be used for prevention and prediction of blood fat. In other words, knowledge discovery from data can help clinicians to predict the future behavior of patients. Also, in line with the development of the system and collect detailed information from patients Such as the psychological state of individuals, Employment

conditions, lifestyle, stress and selection of the health centers in different data sets more stringent rules can be created to predict. To detect the high-risk individuals, Medium and low risk of disease patients at an early stage and changes in lifestyle and follow a disease risk factors, helped to prevent the disease.

REFERENCES

[1] Rahimi shateranlo, E. And Alizadeh, S., 2014. Predict coronary heart disease using a combination of data mining models. Iranian conference: soft computing and IT 2014, Volume-3 Issue-1

[2] Safdari, R., Ghazi.s, M., Gharooni, M., Nasiri, M. And Arji, G.,2014. Compare the performance of decision trees and neural network in the prediction of myocardial infarction. Journal of Mashhad Medical Sciences and Rehabilitation,Volume-3 Issue-1

[3] Zamanpoor, S. And Shamsi, M., 2012. Comparative evaluation of the accuracy of data mining algorithms to predict heart disease, Fourth Conference Electrical and Electronic Engineering,iran-gonabad.

[4] Kashefi.k, A., Pormousa, A., and jahanbani, A.,2007. Multi-layer neural network training using the PSO algorithm, Eighth Conference Intelligent Systems,mashhad-iran.

[5] Crawford M., 2009. Current diagnosis & treatment in cardiology 2009. 3rd ed. Newyork: mcgraw-Hill Medical.

[6] Mobley, B. A., Schechter, E., Moore, W. E., mckee, P. A., and Eichner, J. E. (2005). Neural network predictions of significant coronary artery stenosis in men. Artificial intelligence in medicine, 34(2), 151-161.

[7] Nahar, J., Imam, T., Tickle, K. S., and Chen, Y. P. P. (2013). Association rule mining to detect factors which contribute to heart disease in males and females. Expert Systems with Applications, 40(4), 1086-1093.

[8] Bennetts, C. J., Owings, T. M., Erdemir, A., Botek, G. And Cavanagh, P. R. (2013). Clustering and classification of regional peak plantar pressures of diabetic feet. Journal of biomechanics, 46(1), 19-25.

[9] Canivell, S. And Gomis, R. (2014). Diagnosis and classification of autoimmune diabetes mellitus. Autoimmunity reviews, 13(4), 403-407.

[10] Ordon, M., Urbach, D., Mamdani, M., Saskin, R., Honey, R. J. D. A. And Pace, K. T. (2014). The surgical management of kidney stone disease: a population based time series analysis. The Journal of urology, 192(5), 1450-1456.

[11] Amato, F., López, A., Peña-Méndez, E. M., Vaňhara, P., Hampl, A. And Havel, J. (2013). Artificial neural networks in medical diagnosis. Journal of applied biomedicine, 11(2), 47-58.

[12] Santhanam, T. And Padmavathi, M. S. (2015). Application of K-Means and Genetic Algorithms for Dimension Reduction by Integrating SVM for Diabetes Diagnosis. Procedia Computer Science, 47, 76-83.

[13] López-Chau, A., Cervantes, J., López-García, L. And Lamont, F. G. (2013). Fisher's decision tree. Expert Systems with Applications, 40(16), 6283-6291.

[14] Lappenschaar, M., Hommersom, A., Lucas, P. J., Lagro, J. And Visscher, S. (2013). Multilevel Bayesian networks for the analysis of hierarchical health care data. Artificial intelligence in medicine, 57(3), 171-183.

[15] Han, J., Kamber, M. And Pei, J. (2011). Data mining: concepts and techniques: concepts and techniques. Www.Elsevier.com

[16] Ezanjani, H. Introduction to data mining, www.hajarian.com/IT/tahghigh/zanjani.pdf

[17] Rezai, A., Keshavarzi, P., and Mahdiye, R. (2014). A novel MLP network implementation in CMOL technology. Engineering Science and Technology, an International Journal, 17(3), 165-172.

[18] Wang, C., Li, L., Wang, L., Ping, Z., Flory, M. T., Wang, G., and Li, W. (2013). Evaluating the risk of type 2 diabetes mellitus using artificial neural network: An effective classification approach. Diabetes research and clinical practice, 100(1), 111-118.

[19] Saritha, M., Joseph, K. P., & Mathew, A. T. (2013). Classification of MRI brain images using combined wavelet entropy based spider web plots and probabilistic neural network. Pattern Recognition Letters, 34(16), 2151-2156.

[20] Jalalian, A., Mashohor, S. B., Mahmud, H. R., Saripan, M. I. B., Ramli, A. R. B., & Karasfi, B. (2013). Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. Clinical imaging, 37(3), 420-426.

[21] Bala, S., & Kumar, K. (2014). A Literature Review on Kidney Disease Prediction using Data Mining Classification Technique.

[22] Bajaj, P., Choudhary, K., &Chauhan, R. (2015). Prediction of Occurrence of Heart Disease and Its Dependability on RCT Using Data Mining Techniques. Ininformation Systems Design and Intelligent Applications (pp. 851-858). Springer India.

[23] Suykens, J. A. And Vandewalle, J. (1999). Least squares support vector machine classifiers. Neural processing letters, 9(3), 293-300.

[24] Basak, D., Pal, S. And Patranabis, D. C. (2007). Support vector regression.Neural Information Processing-Letters and Reviews, 11(10), 203-224.

[25] Fadini, G. P. And Avogaro, A. (2013). Diabetes impairs mobilization of stem cells for the treatment of cardiovascular disease: a meta-regression analysis. International journal of cardiology, 168(2), 892-897.

[26] D'Ascenzo, F., Agostoni, P., Abbate, A., Castagno, D., Lipinski, M. J., Vetrovec, G. W., ... And Gaita, F. (2013). Atherosclerotic coronary plaque regression and the risk of adverse cardiovascular events: a meta-regression of randomized clinical trials. Atherosclerosis, 226(1), 178-185.

[27] Soni, J., Ansari, U., and Shrma, D. 2010. Intelligent and Effective Heart Disease Prediction System using Weighted Associative Classifiers, IJCSE.

[28] Mohammadpour Tahamtan, A., Esmaeili, M., Ghaemian, A. And Esmaeili.J.2012. Application of Artificial

Neural Network for Assessing Coronary Artery Disease,J Mazand Univ Med Sci, 2012, 22(86) 9-17.

[29] Jyoti, S., Ujma, A., Dipesh, S. And Sunita, S. 2011. Predictive Data Mining for Medical Diagnosis. An Overview of Heart Disease Prediction, International Journal of Computer Applications 2011, 17(8): 35-43.

[30] Biglarian, A., Babaee, R. And Azmie, R. 2004. Application of Artificial Neural Network Model in Determining Important Predictors of In Hospital Mortality After Coronary Artery Bypass Graft Surgery, and it's Comparison with Logistic Regression Model ,Modarres J Med Sci 2004, 7(1), 23-30. [Persian]

[31] Colombet, I., Ruelland, A., Chatellier, G., Gueyffier F., Degoulet, P. And Christine, M. 2000. Models to predict cardiovascular risk: comparison of CART, Multilayer perception and logistic regression. Proc AMIA Symp 2000:156-160.

[32] Dubey, A., Patel, R. And Choure, K. 2014. An Efficient Data Mining and Ant Colony Optimization technique (DMACO) for Heart Disease Prediction, International Journal of Advanced Technology and Engineering Exploration, Volume-1 Issue-1 December-2014.

[33] Fadini, G. P. And Avogaro, A. (2013). Diabetes impairs mobilization of stem cells for the treatment of cardiovascular disease: a meta-regression analysis. International journal of cardiology, 168(2), 892-897.

[34] Chau, K.W. and Cheng, C.T., 2002. Real-time prediction of water stage with artificial neural network approach. Lecture Notes in Artificial Intelligence 2557, 715.

[35] Rumelhart, D.E., Hinton, E. And Williams, J., 1986. Learning internal representation by error propagation. Parallel Distributed Processing 1, 318–362.

[36] Bazartseren, B., Hildebrandt, G., Holz, K.-P., 2003. Short-term water level prediction using neural networks and neuro-fuzzy approach. Neurocomputing 55 (3–4), 439–450.

[37] Haykin, S., 1999. Neural Networks, A Comprehensive Foundation. Prentice Hall, Upper Saddle River.

[38] Rogers, L.L., Dowla, F.U. and Johnson, V.M., 1995. Optimal field-scale groundwater remediation using neural networks and the genetic algorithm. Environmental Science and Technology 29 (5), 1145– 1155.

[39] Rumelhart, D.E., Hinton, E. And Williams, J., 1986. Learning internal representation by error propagation. Parallel Distributed Processing 1, 318– 362.

[40] Clerc, M. And Kennedy, J., 2002. The particle swarm-explosion, stability, and convergence in a multidimensional complex space. EEE Transactions on Evolutionary Computation 6 (1), 58–73.

[41] Konstantinos E. Parsopoulos and Michael N. Vrahatis, 2004. On the Computation of All Global Minimizers Through Particle Swarm Optimization, IEEE transactions on evolutionary computation, vol. 8, no. 3, june 2004