

MYOCARDIAL INFARCTION DETECTION USING INTELLIGENT ALGORITHMS

V. Seenivasagam*, R. Chitra†

Abstract: Myocardial Infarction (MI) also known as heart attack is one of the most dangerous cardiovascular diseases. Accurate early prediction can effectively reduce the mortality rate caused by MI. The early stages of MI may only have subtle indications which can be varied in variable risk factors and making diagnosis difficult even for experienced cardiologists. In this paper the computer aided detection system is proposed to find the risk level of MI using the supervised classifier. The MI prediction system is developed using Feed Forward Neural Network (FFNN), Cascade Correlation Neural Network (CNN), and Support Vector Machine (SVM). Genetic Optimized Neural Network (GAANN), Particle Swarm Optimized Neural Network (PSO) and also the performance of the Computer Aided Detection system is analyzed using various performance metrics.

Key words: *intelligent algorithms, risk factors, myocardial infarction, feed forward neural network, computer aided detection*

Received: April 21, 2014

DOI: 10.14311/NNW.2016.26.005

Revised and accepted: January 29, 2015

1. Introduction

Ischemic heart disease is the most common type of cardiovascular disease around the world. It refers to problems with the circulation of blood to the heart muscle. A partial blockage of one or more of the coronary arteries can result in a lack of enough oxygenated blood. A complete blockage of an artery causes damage to the tissues or a myocardial infarction is commonly known as a heart attack. Ischemia often causes chest pain or discomfort known as angina pectoris [20]. Knowing the early warning signs of heart attack is critical for prompt recognition and treatment. A person experiencing a heart attack may not even be sure of what is happening [27]. Heart attack symptoms vary among individuals and victims of MI may experience diversity of the symptom. Hence, early prediction of MI plays a major role in the diagnoses of disease. Through the early recognition of symptoms one can get the most effective clinical treatment for the best outcome [10]. As with many labor-intensive occupations, cardiologists use Computer Aided Detection (CAD)

*Vellachamy Seenivasagam, Department of Computer Science and Engineering National Engineering College, Kovilpatti, India, E-mail: yespee1094@yahoo.com

†Retnaswamy Chitra – Corresponding author, Department of Computer Science and Engineering Noorul Islam Centre for Higher Education, Kanyakumari District, India, E-mail: jesi_chit@yahoo.co.in

systems that can identify potential MI on heart disease dataset. These software systems are entering clinical practice as a way to improve cardiologist's ability to detect few cases of MI without chest pain. This paper concentrates on developing a CAD system as an artificial second cardiologist using MATLAB software. Classification systems can help in minimizin[g possible errors and also can provide instant examination of medical data in shorter time and in a more detailed manner. New artificial intelligent techniques such as neural network have been used in medical applications for detecting the normal and abnormal patient from dataset. The patient with high risk of MI is classified as abnormal and others are classified as normal. The thriving of artificial intelligence which utilizes the human experience in a more relaxed form than the conventional mathematical approach has recently attracted more attention. The classification is done using supervised classifiers. Three most popular supervised classifiers were used for classification: Feed Forward Neural Network, Cascade Correlation Neural Network and Support Vector Machine. Designing optimal neural network architecture is made by a human expert and it requires a tedious trial and error process. Especially automatic determination of artificial neural network parameters is the most critical task. Soft computing evolutionary approaches like Genetic algorithms and Particle Swarm Intelligence can be effectively used for tuning the intelligent networks. This paper focuses mainly on the designing and analysis of a CAD system based on supervised classifiers and the optimized neural network parameters evaluated using evolutionary approaches to improve the classification accuracy in MI detection thereby reducing the misclassification rate.

The rest of the Section is organized as follows. Section 2 presents a detailed literature survey and the past works involved in the design, application and analysis heart disease detection system using data mining and intelligent algorithm. Section 3 explores the proposed system model and the detailed model used for myocardial infarction detection. The application of supervised classifiers used for detecting the MI is analyzed for Cleveland dataset and real time clinical database is presented in Section 4 which includes the intelligent algorithms like Feed Forward Neural Network, Cascade Correlation Neural Network and Support Vector Machine. The FFNN needs tuning of its parameters for optimal performance. To achieve the optimality in artificial neural network various evolutionary approaches like Genetic Algorithm and Particle Swarm Optimization were used .The various parameters considered for optimization are learning rate, momentum factor and number of neurons in the hidden layer and these paradigms are presented in Section 5. In Section 6, a concise review of work reported, contributions made and comparative performance results for the MI detection is described. The major conclusions reached are presented with the recommendation for future work in Section 7.

2. Related work

Last few decades many intelligent techniques were developed by lot of researches in the area of disease diagnosis. Different studies have demonstrated that Computer Aided Detection of MI can help as a secondary tool for the cardiologist. Regarding classification of MI, a number of techniques have been presented using machine learning approaches to classify samples as normal and abnormal.

Tarek et al. [27] proposed ensemble and hybrid intelligent techniques such as Support Vector Machine, Function Network and Fuzzy Logic to classify bioinformatics datasets. Saangyong et al. [23] have presented the machine learning techniques, SVM, decision tree, and decision rule to predict the susceptibility of the liver disease, chronic hepatitis from single nucleotide polymorphism data. Ozyilmaz and Yildirim [22] have presented three neural network algorithms for diagnosis of hepatitis diseases, multilayer neural network produced significantly good result for the diagnosis. Recently, SVM have emerged as a powerful technique for general purpose pattern recognition and it has been applied to classification and regression problems with exceptionally good performance on a range of binary classification tasks [1, 6, 16]. The primary advantage of SVM is its ability to minimize both structural and empirical risk [11] leading to better generalization for new data classification. Normally, direct support clinical decision making is the intention behind the design of a clinical decision support system and it presents patient specific assessments or recommendations produced using the characteristics of individual patients to clinicians for consideration [15].

Ischemic heart diseases have emerged as the number one killer in most of the countries, in case of heart disease time is very crucial to get correct diagnosis in early stage. An automated system can help to medical community to assist doctor for the accurate diagnosis well in advance [17]. As per a survey nearly 50 percent of patients, however, have no symptoms until a heart attack occurs. A number of factors have been shown to increase the risk of developing MI. They proposed combination of data mining techniques using new clinical dataset many factors to analyze for a diagnosis of heart disease. [29]. Deepika et al. [4] proposed association rule for classification of Heart attack patients. The significant patterns were extracted from the heart disease warehouse and association rule is used to classify the normal and abnormal patterns. The frequent item set mining was used and the significant items were chosen and it was used in the design and development of the heart attack prediction system. Shanthakumar et al. [25] introduced K-means clustering algorithm to extract the data appropriate to heart attack from the warehouse. In addition the pattern vital to heart attack were selected on basis of the computer significant weightage.

Srinivasan et al. [26] proposed application of data mining technique in healthcare and prediction of heart attacks. Rule based, Decision tree and Naive Bayes were applied to healthcare data classification decision making concept with Naive Bayes classifier was explained. Association rule mining based on heart attack prediction was proposed by Jabbar et al. [14]. The dataset with 13 input attributes was used in that work and also mining frequent item sets was proposed with reduced main memory requirement and space complexity was reduced. Chaitrali et al. [2] discussed the improved data mining techniques for heart attack prediction. Different classification techniques were used for extracting significant pattern from heart disease dataset to improve classification accuracy was proposed by Sellappan et al. [24]. Many Researches used UCI machine learning repository database for experimental verification [21]. CNN has been proved to be a well-known decision support system for prediction in nonlinear applications [5]. The result illustrated the peculiar strength of each of the methodologies in comprehending the objectives of the specified mining objectives. It facilitated the establishment of vital knowledge, e.g. patterns connected with heart disease.

To utilize artificial intelligence tools as clinical decision support in assessing myocardial risk in patients, Fidele et al. [9] presented a research study. In the proposed artificial neural network, a two-layer neural network employing the Levenberg-Marquardt algorithm and the resilient back-propagation have been utilized. The integration of GAANN in medical domain improves performance and accuracy [7], improved prediction rate [12] and optimize the number of neurons in hidden layer [13]. Few models that integrate FFNN with GA have been reported in medical literature for diagnosis such as diagnosis of critically ill [19], improved classification performance. Chitra and Seenivasagam proposed genetic optimized FFNN for heart disease prediction system; in their work they proved significant improvement in performance using genetic optimization [3]. Kuok et al. proposed neural network trained with back propagation algorithm to calibrate the rainfall-runoff relationship accurately using only rainfall and runoff data [18]. They introduced, a new evolutionary algorithm namely Particle swarm optimization to train the feed forward neural network and thus the local minima convergence problem in backpropagation learning was eliminated. This Particle Swarm Optimization Feed forward Neural Network was applied to model the hourly rainfall-runoff relationship for Bedup Basin. With the input data of current rainfall, antecedent rainfall, antecedent runoff, the optimal configuration of PSONN successfully simulate current run off accurately.

Among existing techniques, the main problem of developing an acceptable CAD system is inconsistent and low classification accuracy. In order to improve the training process and accuracy, this work investigates novel intelligent classifiers SVM, CNN and FFNN that use data attributes as input to classify the normal and abnormal. Moreover, the intelligent machine learning classifier FFNN is optimized using heuristic algorithms for finding appropriate hidden neurons, learning rate and momentum constant during the training process. From the literature it is proved that GA and PSO are the suitable algorithms for FFNN optimization.

3. Proposed system model

MI detection using a CAD system is an extremely challenging task. First, since CAD systems are computer directed system, there is a need for a flawless system. Second, the large variability in the identification of abnormal patient with one or two symptoms. Finally, the patient without external symptom can also have the possibility of MI. Hence, there is a need for analyze the attributes of the MI dataset and an intelligent classifier to classify the abnormalities among the data.

The proposed CAD system is based on a data processing system which intelligently classifies the high risk patients as abnormal. CAD schemes using digital data processing techniques have the goal of improving the detection performance. Typically CAD systems are designed to provide a second opinion to aid rather than replacing the cardiologist. The proposed system has been designed in a framework of MATLAB 7.10, which aims at developing an intelligent system for MI detection. The proposed system architecture is shown in Fig. 1.

MI dataset is obtained from UCI (University of California, Irvine C.A) centre for machine learning and intelligent systems named Cleveland dataset (270 records) and from various cardio labs named Clinical dataset (500). In this datasets first

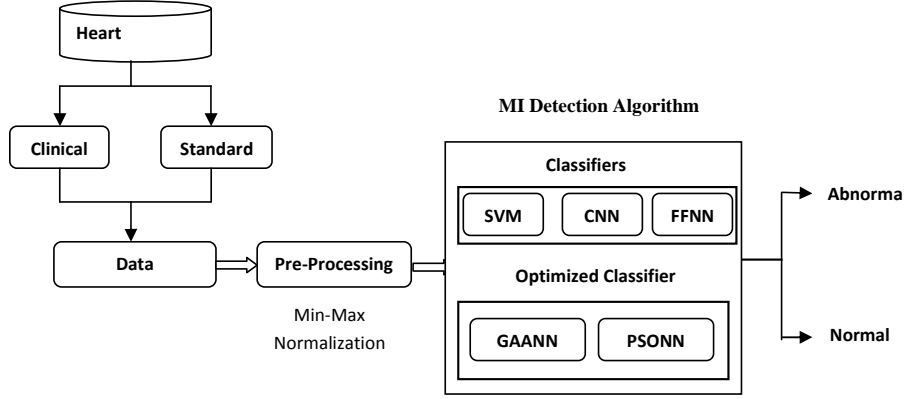


Fig. 1 *Proposed intelligent MI detection system.*

13 attributes describes the risk factors of ischemic heart disease and last attribute describes the output class. There are two output classes for the diagnosis of heart attack. In the selected dataset, class 0 specifies without the risk of MI and class 1 specifies the risk of MI.

In the proposed system the goal of pre-processing is to simplify recognition of patient with high risk factor without throwing away any important information. The normalization technique used in intelligent MI prediction is min-max normalization. Let A be a numeric attribute with n observed values, v_1, v_2, \dots, v_n . Min-max normalization performs a linear transformation on the original data. Suppose that \min_A and \max_A are the minimum and maximum value of an attribute A . Min-max normalization maps a value v_i of A to v'_i in the range new - \min_A , new - \max_A by computing

$$v'_i = \frac{v_i - \min_A}{\max_A - \min_A} (\text{new} \dots \max_A - \text{new} \dots \min_A) + \text{new} \dots \min_A. \quad (1)$$

Min-max normalization preserves the relationship among the original data values. The values are normalized between the limit $[0, 1]$ and then the Eq. (1) becomes

$$v'_i = \frac{v_i - \min_A}{\max_A - \min_A}. \quad (2)$$

In the proposed system supervised classifiers are used to separate normal and abnormal patients. The ultimate objective of automated methods for classification of abnormalities in heart disease dataset is to provide a tentative diagnosis and the final decision about the MI produced by a human expert based on their physical attributes. The performance of classification methods depends on the type and quality of the features employed to train the classifier. Three most popular classifiers are used for classification: Feed forward neural network, Cascaded correlation neural network and Support Vector machine. Soft computing optimization algorithms are used to optimize the hidden layer neurons, learning rate and momentum constant of FFNN to increase the accuracy of the classifier. The random population-based algorithms PSO and GA are used for optimization.

4. Artificially intelligent approaches for MI detection

The intelligent MI detection system with SVM, CNN and FFNN is explained in this Section. It includes the general concept of the specified intelligent classifier and how it is suitable for detecting MI.

4.1 Myocardial infarction detection using SVM

Support Vector Machine is used as an effective computational intelligent technique to classify disease from large bioinformatics datasets. SVM is a set of related supervised learning method used in medical diagnosis for classification and regression. A SVM performs classification by constructing a high dimensional hyper plane that optimally separates the data into two categories. SVM models are closely related to neural networks. In fact, a SVM model using a sigmoid kernel function is equivalent to a two-layer, perceptron neural network. A predictor variable is called an attribute, and a transformed attribute that is used to define the hyper plane is called a feature. The task of choosing the most suitable representation is known as feature selection. A set of features that describes one case (i.e., a row of predictor values) is called a vector. So the goal of SVM modeling is to find the optimal hyper plane that separates clusters of vector in such a way that cases with one category of the target variable are on one side of the plane and cases with the other category are on the other side of the plane.

Consider an n -dimensional pattern (object) x has n coordinates $x = \{x_1, x_2, \dots, x_n\}$, where x_i is a real number and $x_i \in R$, for $i = 1, 2, \dots, n$. Each pattern x_j belongs to a class $y_j \in \{-1, +1\}$. The output is grouped as normal if $y_j = -1$ and abnormal if $y_j = +1$. Consider a training set T of m patterns together with their classes.

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}. \quad (3)$$

The patterns are embedded into a dot product space, $x_1, x_2, \dots, x_m \in S$. Any hyper plane in the space can be written as

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}. \quad (4)$$

The patterns are embedded into a dot product space S , $x_1, x_2, \dots, x_m \in S$. Any hyper plane in the space S can be written as

$$(x \in S | w \cdot x + b = 0), \quad (5)$$

$$wx = \sum_{i=1}^n w_i x_i. \quad (6)$$

The linear classifier is defined as $(w \cdot x + b = 0)$ and defines a region for class $+1$ as $(w \cdot x + b > 0)$ and -1 as $(w \cdot x + b < 0)$. After training the classifier is ready for prediction of new patterns, different from those used in training. For nonlinear classifier models, SVM maps the input vector into a higher dimensional feature space \mathcal{H} and an optimal separating hyper plane is constructed. A nonlinear transformation $\varphi(x)$ is done for mapping, using the suitable basis function and then uses the

linear model in the feature space, which is called the kernel function. The SVM classification is written in the following form:

$$f_{\text{svm}}(x) = w^T \varphi(x) + b, \quad (7)$$

where the parameter w is the weight factor and b is the bias, which are determined from training data set. SVM finds the support vectors that define the widest separation of classes. The merit of support vector machine is that, by a kernel function $K(x_i, y_i)$, which is the inner product in the feature space, it tries to make the training data linear separability in the high dimension feature space, thus achieve non-linear separability in the input space. A training sample (x_i, y_i) is a support vector if it holds $y_i w^T \varphi(x) + b \geq 1$. Let S_k be the support vector, $k \in [1, K]$, and the SVM function becomes

$$f_{\text{SVM}}(x) = \sum_{k=1}^K K(x, S_k) + b, \quad (8)$$

$$K(x, S_k) = \varphi^T(x) \varphi(S_k). \quad (9)$$

$K(.,.)$ represents the kernel function to represent the effect of non-linear mapping and $\varphi(.)$ in classification of MI. A radial basis function network is used as a kernel function for classification of MI, which maps the input space to a higher dimensional feature vector and is given as

$$K(x, S_k) = \exp \frac{-||x_i - S_k||^2}{2\sigma^2}. \quad (10)$$

The strategy is to map the training data to the higher dimensional feature space and include most of them into a hyper sphere of minimum size. That is, the task is to minimize the following objective function

$$\text{Minimize } \frac{||w||^2}{2} + c \sum_{i=1}^n e_i, \quad (11)$$

where c is the regularization parameter controlling the tradeoff between margin maximization and classification error. E is called the slack variable that is related to classification errors in SVM. RBF kernel function can analyze higher dimensional data. The output of the kernel is dependent on the Euclidean distance of x_j from x_i (one of these will be the support vector and the other will be the testing data point). In this work RBF kernel function is used because RBF kernel nonlinearly maps samples into a higher dimensional space and also it has less numerical difficulties.

The performance of the system was analyzed with 270 data in Cleveland and 500 data in clinical data. The dataset with 13 attributes is fed to the SVM classifier and the output is mapped into two classes normal and abnormal. The choice of best value of parameters for particular kernel is critical for a given amount of data. The features are classified into two classes via abnormal and normal. Abnormal samples are collected (MI present) and assigned a value +1 and normal samples are assigned a value -1 (MI absent). Once the training examples are athered, the next step is to determine the SVM decision function. The training set completed in

53 iterations with a threshold fixed at 0.01. These features are trained by the SVM classifier and the classification power of these features is tested using the test cases. For automatic classification of abnormalities the 13 attributes were used, this set of features provides a ground for computer scheme to achieve high classification performance. These attributes are treated as an input pattern and is labelled as +1 for high risk of MI patient and -1 for low and absent of MI risk. Together (x, y) forms an input-output pair, where x is the input feature and y is the output (abnormal/normal).

4.2 Myocardial infarction detection using CNN

The cascaded correlation network uses a hierarchical arrangement of hidden units and it is a special type of neural network. A cascade correlation network consists of a cascade architecture, in which hidden neurons are added to the network one at a time and do not change after they have been added. The architecture learns very quickly since it requires no back propagation of error signals through the connections of the network. A cascade correlation network consists of a cascade architecture, in which hidden neurons are added to the network one at a time and do not change after they have been added. It is called a cascade because the output from all neurons already in the network feed into new neurons. As new neurons are added to the hidden layer, the learning algorithm attempts to maximize the magnitude of the correlation between the new neurons output and the residual error of the network which we are trying to minimize. In a CNN, the number of input nodes is determined by the input features, while the number of output nodes is determined by the number of different output classes. Initially the network contains only inputs, output units, and the connections between them. This single layer of connections is trained using the Quick propagation algorithm [8] to minimize the error. Cascade correlation eliminates the need for the user to guess in advance the network's size, depth, and topology; it uses simple training rules since only one layer of weights is being trained at a time. A cascade correlation neural network has three layers: input unit, hidden unit and output unit. The network begins with an input and the output unit and no hidden units. The number of inputs I_p and the output O_p is defined by the problem, which is defined as $I_p = \{i_1, i_2, \dots, i_p\}$ and $O_p = \{o_1, o_2, \dots, o_p\}$ respectively.

Every input is connected to the output unit with the weights obtained from the BP algorithm. There is also a bias which is set to +1. The hidden units are added to the network one by one until the error is minimized or the stopping criterion is reached. Each new hidden unit receives a connection from each of the network's original inputs and also from every pre existing hidden unit. To create a new hidden unit, to satisfy the condition that $S = \sum_o \left| \sum_p (V_p - \bar{V}) (E_{p,o} - \bar{E}_o) \right|$ should be maximized, where o is the output unit, p is a number of patterns in the training set, V_p is the candidate units value at p , $E_{p,o}$ is the residual error of all the training pattern at the output unit. \bar{V} and \bar{E}_o are the values of V and E_o averaged over all patterns. The residual error is calculated using the following equation:

$$E_{p,o} = (Y_{p,o} - T_{p,o}), \quad (12)$$

where $Y_{p,o}$ is the actual output at the output unit o , $T_{p,o}$ is the desired output at the output unit o . The training is done using the back propagation algorithm. Backdrop algorithm uses a gradient descent method to update the weights. The steps involved in back propagation algorithm are:

Step1: Initialize the input and output units based on the problem defined. The input and the Output neurons are fully connected.

Step 2: Train the network with input and output neurons until the residual error $E_{p,o}$ no longer decreases.

Step 3: Select a temporary unit (Candidate unit) connected with the input unit and find the residual error.

Step 4: Train this network unit $S = \sum_o \left| \sum_p (V_p - \bar{V}) (E_{p,o} - \bar{E}_o) \right|$ no longer improves.

Step 5: Connect the temporary unit with the output unit and freeze its weights.

Step 6: Train the input, output and the hidden unit until the residual error is minimized.

Step 7: Repeat the Step 2 to Step 6 until the net error falls below a given value.

In the proposed CAD system CNN classifier is designed with 13 input neurons, one hidden layer and one neuron at the output layer and it is shown in Fig. 2. The output layer classifies the given input feature as normal and abnormal pattern. The hidden layer neuron is optimally chosen by the CNN network. The hidden layer uses a sigmoid activation function. The learning rate and momentum constant was set to 0.01 and 0.9 respectively. The training was stopped when the root mean square error per training was less than 0.1 or when it reaches 500 epochs. After training the network was evaluated with the test cases. Fig. 2 shows the simulated structure of CNN used for MI prediction. The w is the weights connecting the neurons and b is the bias. The input layer neurons are connected to the hidden layer neurons and also the output layer neurons. This brings a cascaded structure to the neural network.

4.3 Myocardial infarction detection using FFNN

FFNN is a network having nonparametric statistical, it can learn the free parameters weights and biases through training by using examples. Hence it can be used for nonlinear and complex problems. The synapses of the biological neuron are modelled as weights. The weights are the one which interconnects the neural network and gives the strength of the connection. All the inputs are summed together and are modified by the weights. Let w_1, w_2, \dots, w_n are weights to determine the strength of input vector $X = [x_1, x_2, \dots, x_n]$. Each input is multiplied by the associated weight of the neuron connection $X^T w$. The positive weight excites and the negative weight inhibits the node output.

$$I = X^T w = x_1 w_1 + x_2 w_2 + \dots + x_n w_n = \sum_{i=0}^n x_i w_i. \quad (13)$$

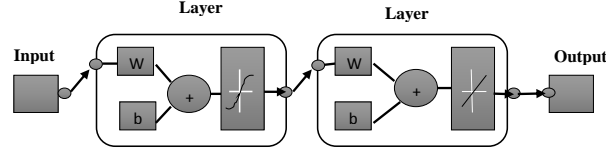


Fig. 2 *Proposed design of CNN.*

In practice, neurons generally do not produce an output unless their total input goes above a threshold value. The neuron output is the sum of the weighted inputs to the neuron minus the threshold value and is given as

$$y = f[ERR : md : MbegChr = 0x007B, MendChr = 0x0029, nParams = 1] - P\theta_k, \quad (14)$$

where θP_k is the threshold value and b is bias. The final output y is generated by passing the sum on to a nonlinear filter f called the activation function. The activation function f performs mathematical operation on the output. The Back Propagation (BP) algorithm [9] is used in FFNN training.

The idea of the BP algorithm is to reduce this error, until the FFNN learns the training data. The training begins with random weights, and the goal is to adjust them so that the error will be minimal. In machine learning, one often seeks to predict an output variable y based on a vector x of input variables. To accomplish this, it is assumed that the input and output approximately obey a functional relationship $y = f(x)$, called the predictive model. In supervised learning, the predictive model is discovered with the benefit of training data consisting of examples for which both x and y is known. We will denote these available pairs of examples as (x_i, y_i) , where $i = 1, 2, d \dots n$, and we will assume that x is composed of n variables. The weight coefficient reflects the degree of importance of the given connection in the neural network. The Sigmoid activation function $f(x) = \frac{1}{1+e^{-x}}$ is used to train the patterns. The network uses a bias (+1), a sigmoid function for hidden layer, and a linear function for the output layer for classifying the input vectors into abnormal and normal patient.

Mean square error function E is used in the neural network training and it can be defined as $E = \frac{1}{n} \sum_{i=1}^n (y_i - t_i)^2$. The gradient of the error function is computed and used to correct the initial weights, so that the error function gets minimized. The weights are adjusted using the gradient descent, $\Delta w_{ij}(t+1) = -\eta \frac{\partial E}{\partial w_{ij}} + \alpha \Delta w_{ij}(t)$, where α is the momentum factor and η is the learning rate. The momentum factor α allows for momentum in weight adjustment. Momentum basically allows a change to the weights to persist for a number of adjustment cycles. The magnitude of the persistence is controlled by the momentum factor. Increased momentum factor increases greater persistence of previous adjustments in modifying the current adjustment. Momentum in the back propagation algorithm can be helpful in speeding the convergence and avoiding local minima. Learning rate η is a training parameter that controls the size of weight and bias changes during learning the selection of a distance.

The features were captured with different kind of attributes and a total of 13 features were selected. FFNN classifier is thus designed with 13 input neurons, one

hidden layer with 6 neurons and one neuron at the output layer. The output layer classifies the given input feature as normal and abnormal pattern. The learning rate and momentum constant was set to 0.01 and 0.9 respectively. The training was stopped when the root mean square error per training was less than 0.1 or when it reaches 500 epochs. After training, the network was evaluated with the test cases. The simulated structure of FFNN used in this work is shown in Fig. 3.

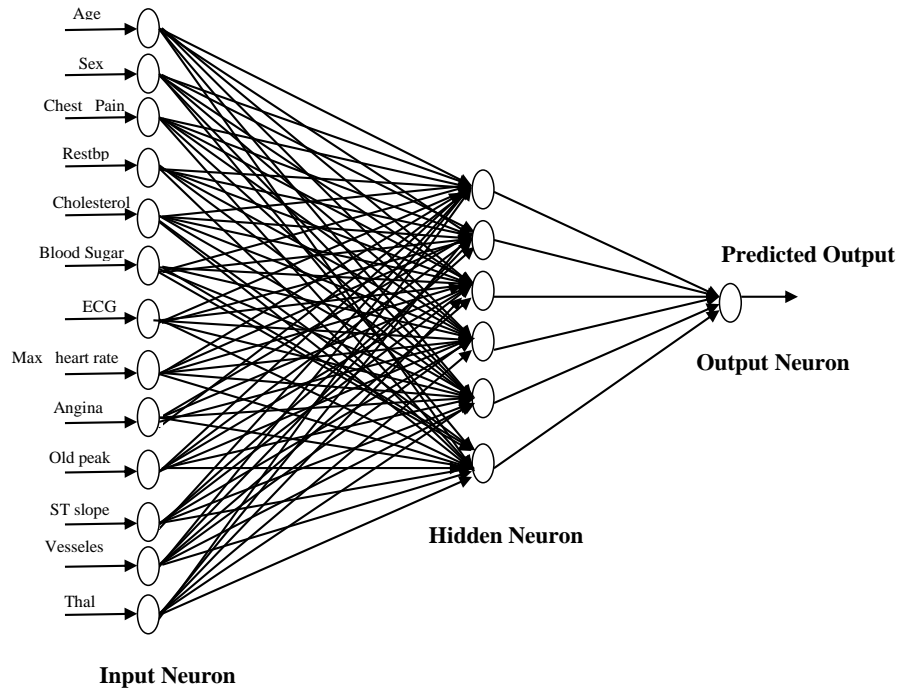


Fig. 3 Simulated structure of FFNN.

5. Optimized neural network based MI detection

To improve the prediction accuracy of MI detection heuristic algorithms genetic algorithm and particle swarm optimization are used and are explained in this Section.

5.1 Genetic optimized FFNN based MI detection

Genetic algorithms are computerized search and optimization algorithms based on the mechanics of natural genetics and natural selection. The main idea there is to form a group of alternative solutions for the problem, that is, a population, and then to breed this initial population according to evaluative laws. A fitness function is used to measure the quality of each individual and to select the best individuals to produce offspring. GA has been used to search for the optimal neurons in the

hidden layer, connectivity, training parameter such as momentum rate, learning rate for FFNN for predicting MI. For each chromosome of the population fitness considered is the mean square error (MSE) computed by FFNN. Once fitness is calculated for all chromosomes, then it replaces the worst fitness to best fitness value. Further crossover is performed by using two-point crossover with the given number of generation. Finally mutation is applied which is used to generate the new population. The new population can be given to the input to compute the fitness of each chromosome, followed by the same GA operation to generate the next population. This process is repeatedly performed until the all the chromosome gets near to the same fitness value. Genetic Optimization is introduced in the back propagation learning for optimizing the learning rate, hidden layer neurons and momentum constant.

In GAANN method, FFNN is optimized with best network architecture by optimizing the number of neurons in the hidden layer, the learning rate and the momentum factor. Finding an optimal learning rate avoids major disruption of the direction of learning when very unusual pair of training patterns is presented. The number of neurons in the input layer and output layer is fixed based on the problem defined. Let N_I represents the number of the neurons in the input layer and N_O represents the number of the neurons in the output layer. The number of neurons in the input and output layer are fixed and they are same for the entire configuration in the architecture spaces. The number of hidden layers in this problem is restricted and made to one. The range of the optimization process is defined by two range arrays $R_{\min} = \{N_{h\min}, L_{r\min}, M_{c\min}\}$ and $R_{\max} = \{N_{h\max}, L_{r\max}, M_{c\max}\}$, where N_h is the number of neurons in the hidden layer, L_r is the learning rate and M_c is the momentum factor. Let f be the activation function and is defined as the sum of the weighted inputs plus the bias and is represented as $y_k^p = f(s_k^p)$, where $s_k^p = \sum_j w_{j,k} y_j^p + b_k$, y_j^p is the output of the j -th neuron when pattern p is fed, $w_{j,k}$ is the weight from the j -th neuron and b_k is the bias value of the k -th neuron in the hidden layer. The hidden layer uses a sigmoid activation function and the fitness function sought for optimal training is the MSE, it can be formulated as

$$\text{MSE}_{\text{GAANN}} = \sum_{p \in T} \sum_{k=1}^{N_O} (t_k^p - y_k^{p,o})^2, \quad (15)$$

where t_k^p is the desired output, $y_k^{p,o}$ is the actual output from the k -th neuron in the output layer o , for the pattern p in the training set. With the framed fitness function the GAANN algorithm automatically evolve a best solution. Fig. 4 shows the flow diagram of GAANN designs for intelligent MI detection.

The GAANN has three layer architecture, an input layer, hidden layer, and an output layer. The number of neurons that structures the input layer is equal to the number of input attributes that is 13. The hidden layer neurons are optimally added to the FFNN and are defined by the sigmoid activation function. The output layer contains one neuron which discriminates presence of MI and without MI. The neural network architecture space is defined over a multilayer perceptron with the parameters range set as R_{\min} mutation probability is 0.15 and the cross over probability is 0.8. Number of iterations selected for optimization is 500.

The input data to GAANN need to be normalized as FFNN only work with data between 0.001 and 0.999. Using the proposed GAANN algorithm an optimized

FFNN is achieved with $N_h = 15$, $L_r = 0.0101$ and $M_c = 0.922$. Thus, the PSO algorithm yields a compact network configuration in the architecture space rather than the complex ones as long as optimality prevails. The optimized neural network parameters obtained are momentum factor is 0.922, learning rate is 0.0101 and number of hidden neurons is 15.

5.2 Particle swarm optimized FFNN based MI detection

Particle Swarm Optimization is a population-based stochastic search and optimization process. In the basic PSO algorithm the system is initialized with a population of random solutions and searches for optima by updating positions and velocity. All particles have fitness values which are evaluated by the fitness function to be optimized, and have velocities which direct the flying of the particles.

Each particle is updated after every iteration using two values P^{best} and G^{best} . P^{best} is the personal best value, which indicates the best solution achieved so far (i.e., lowest fitness value) and the G^{best} is the global best solution achieved so far by any particle in the population. In PSO, each particle in the population has a velocity $v_i(t)$, which enables it to fly through the problem space. Therefore, each particle is represented by a position $x_i(t)$, and a velocity vector. Dimensions of position and velocity vectors are defined by the number of decision variables in the optimization problem. Modification of the position of a particle is performed by using its previous position information and its current velocity.

$$v_i(t+1) = wv_i(t) + c_1 \text{rand}_1 (P_i^{\text{best}} - x_i(t)) + c_2 \text{rand}_2 (G_i^{\text{best}} - x_i(t)), \quad (16)$$

$$x_i(t+1) = x_i(t) + v_i(t+1), \quad (17)$$

where (t) is the velocity of particle i at iteration t , $x_i(t)$ is the current position of the particle i at iteration t , P_i^{best} is the personal best of particle i , G_i^{best} is the best position in the neighborhood, rand is the random number between 0 and 1, w is the weighting function, c_1 is the cognition learning rate, c_2 is the social learning rate.

Let N_I represents the number of the neurons in the input layer and N_O represents the number of the neurons in the output layer. The number of neurons in the input and output layer are fixed and they are same for the entire configuration in the architecture space. The number of hidden layers in this problem is restricted and made as one. The range of the optimization process is defined by two range arrays $R_{\min} = \{N_{h\min}, L_{r\min}, M_{c\min}\}$ and $R_{\max} = \{N_{h\max}, L_{r\max}, M_{c\max}\}$, where N_h is the number of neurons in the hidden layer, L_r is the learning rate and M_c is the momentum factor. Let f be the activation function and is defined as the sum of the weighted inputs plus the bias and is represented as, $y_k^p = f(s_k^p)$, where $s_k^p = \sum_j w_{j,k} y_j^p + \theta_k$, y_j^p is the output of the j -th neuron when pattern p is fed, $w_{j,k}$ is the weight from the j -th neuron and θ_k is the bias value of the k -th neuron in the hidden layer. The hidden layer uses a sigmoid activation function and the output layer uses linear activation function. The fitness function sought for optimal training is the MSE which is formulated as

$$\text{MSE}_{\text{PSO}} = \sum_{p \in T} \sum_{k=1}^{N_O} (t_k^p - y_k^{p,o})^2, \quad (18)$$

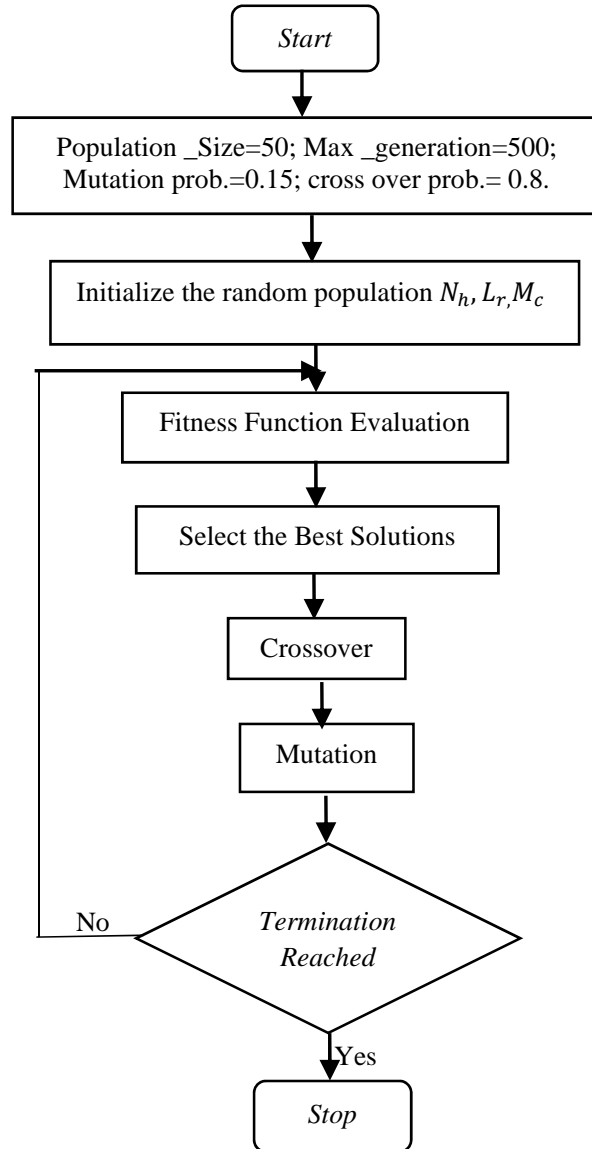


Fig. 4 Flow diagram of GAANN for MI detection.

where t_k^p is the desired output, $y_k^{p,o}$ is the actual output from the k -th neuron in the output layer, for the pattern in the training set. With the framed fitness function the PSNN algorithm automatically evolve a best solution. The optimally designed PSNN has three-layer architecture an input layer with 13 neurons, hidden layer and an output layer. The output layer contains one neuron which discriminates presence of MI and without MI. The neural network architecture space is defined over a multi-layer perceptron with the parameters range set as R_{\min} and R_{\max} .

The PSO algorithm for intelligent MI detection

```

For each particle
    Initialize particle for NN problem
End
Do
    For each particle
        Calculate fitness value (feed forward error or MSE)
        If the fitness value is better than the best fitness value ( $P^{\text{best}}$ ) in history
            Then set current value as the new  $P^{\text{best}}$ 
        End
        Choose the particle with the best fitness value of all the particles as
            the  $G^{\text{best}}$ 
        For each particle
            Calculate particle velocity using Eq. (16)
            Update particle position using Eq. (17)
    End

```

The PSO parameters and c_1, c_2 are the acceleration constants which are initially set to 2 and r_1 and r_2 are random integers set to a range between [0,1]. The initial population is randomly selected and the population size (i.e. the initial set of particles) is set to 25. The total number of iterations (i.e. the maximum generation) is chosen as 200. With these parameters the number of hidden layer neurons, momentum rate and the learning rate, the transfer function in the hidden layer and the learning algorithm are optimized. The whole process is repeated until the generation reaches 200 or the minimum fitness is achieved. During each PSO iteration, the best fitness score (minimum MSE) achieved by the particle with the G^{best} at the optimum dimension is stored. Using the proposed PSO algorithm an optimized FFNN is achieved with $N_h = 19$, $L_r = 0.0112$ and $M_c = 0.925$. Thus, the PSO algorithm yields a compact network configuration in the architecture space rather than the complex ones as long as optimality prevails.

6. Result analysis

Models for prognostic risk prediction have been widely used in the cardiovascular field to predict risk of future events or to stratify apparently healthy individuals into risk categories. The proposed intelligent algorithms are evaluated using two databases Cleveland database and the Clinical database. The Cleveland database contains 270 patient records and the Clinical database contains a set of 500 records collected from various centers. The selected supervised classifiers are trained to discriminate between the abnormal and the normal patient. For evaluation of detection performance, the number of True Positives (TP), False Positives (FP), True negatives (TN) and False Negatives (FN) should be taken into consideration. All findings outside the truth marking by the cardiologists are considered as false detections. In this way the true positive rate can be plotted against the false positive rate. In the proposed MI prediction system obtained results are evaluated by the performance metrics sensitivity, specificity, and accuracy, Sensitivity, specificity and accuracy are the commonly used statistical measures to analyze the medical

diagnostic test to enumerate how the test was good and consistent. The overall performance of diagnostic systems has been measured and reported in terms of classification accuracy which is the percentage of diagnostic decisions that proved to be correct. The various performance metrics for MI prediction using standard Cleveland dataset is shown in Tab. I.

DataSet	Cleveland				
Classifier/Metrics	SVM	CNN	FFNN	GAANN	PSOANN
Accuracy [%]	82.021	85.018	85.41	88.501	89.61
Sensitivity [%]	86.002	87.012	86.67	89.902	90.13
Specificity [%]	77.510	83.031	83.33	86.704	88.98
Youden's index J	0.635	0.7	0.7	0.766	0.79
Miss classification rate	0.1797	0.1498	0.14	0.1149	0.1039

Tab. I *Performance metrics for Cleveland dataset.*

Youden's index is a single statistic that captures the performance of a diagnostic test and is calculated as $J = \text{Sensitivity} + \text{Specificity} - 1$. There are possibilities for an automated classifier to miss certain abnormalities and may miss predict the normal cases as abnormalities, this is known as the misclassification rate and is given as

$$\text{Miss classification rate} = \frac{\text{FP} + \text{FN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}}.$$

Miss classification rate should be less for the good classifier and in the proposed CAD system misclassification rate is significantly acceptable for all classifiers and for PSOANN it is very less hence the classification rate of PSOANN is high compared to SVM, CNN, FFNN, and GAANN. Similarly the performance metrics for Clinical dataset is shown in Tab. II. The PSOANN classifier has less miss classification rate compared to SVM, CNN, FFNN and GAANN for Clinical dataset also.

Dataset	Clinical				
Classifier/Metrics	SVM	CNN	FFNN	GAANN	PSOANN
Accuracy [%]	79.02	80.78	81.21	85.32	85.62
Sensitivity [%]	80.51	82.7	83.67	86.41	88.53
Specificity [%]	77.51	78.6	78.73	83.52	82.67
Youden's index J	0.58	0.613	0.624	0.699	0.71
Miss classification rate	0.2098	0.1922	0.188	0.1468	0.1438

Tab. II *Performance metrics for clinical dataset.*

The Accuracy, Sensitivity and Specificity of the MI prediction system are analyzed using Fig. 5 and 6.

The PSOANN approach generates a sensitivity of 90.13% with a misclassification rate of 0.1039 for Cleveland and a sensitivity of 88.53% with a misclassification rate

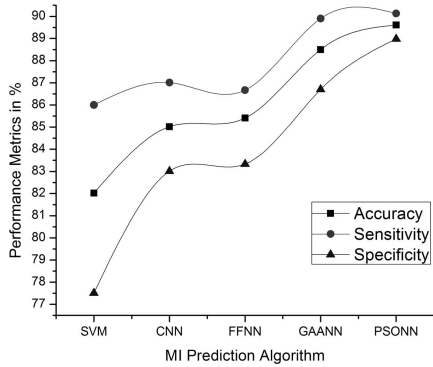


Fig. 5 Performance analysis of Cleveland dataset.

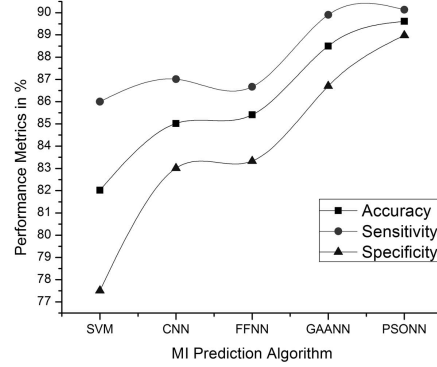


Fig. 6 Performance analysis of Clinical dataset.

of 0.1438 for real time clinical database. When optimized learning is introduced, there is an improvement in Youden's index also. The overall accuracy of the CNN method is found to be 85.6, while the accuracy of the FFNN 81.2 and SVM is 79 for real time clinical dataset containing 500 data. FFNN has an improved performance than the SVM, this is due to the fact that SVM does not depends on the generalization error. Although FFNN produces a good classification accuracy for prediction of heart diseases, this classifier select the initial parameter setting manually which will take a considerable time to classify the patterns. The PSOINN classifier has achieved very commendable results achieving a classification accuracy of 89.61% for Cleveland database and 85.62% for real clinical database when compared with SVM, CNN, FFNN and GAANN. The performance metrics for GAANN is also considerably high compared to FFNN. Accuracy of the system is 85.9%, sensitivity 83.6%, specificity 87.8% for the FFNN based MI risk prediction and Accuracy 88.5%, sensitivity 86.7%, specificity 89.9% for the GAANN based MI risk prediction for standard database. The accuracy of the neural network classifier for heart disease prediction is 0.856. But in the case of Genetic Optimized neural network classifier is 0.885. It is proved that the efficiency of the MI risk prediction system is improved 3% and 4% by optimizing the network parameters using GA and PSO respectively.

An ROC curve is a plot of operating points which can be considered as a plotting of true positive as a function of false positive. ROC curve is a plot of the True Positive Rate (TPR) versus False Positive Rate (FPR). The TPR denote the fraction of patients actually having the abnormality and that are diagnosed as positive and the FPR is the fraction of patients actually without the abnormality and that are diagnosed as positive. The detection performance is analyzed using the area under the ROC curve. For a two class problem an ROC curve allows us to allow us to visualize the trade-off between the rate at which the model. The ROC curve for Cleveland and clinical dataset for different intelligent classifier are shown in Fig. 7 and Fig. 8 respectively.

From this curve it is evident that the feed forward neural network optimized with particle swarm optimization algorithm has found to have better accuracy than the

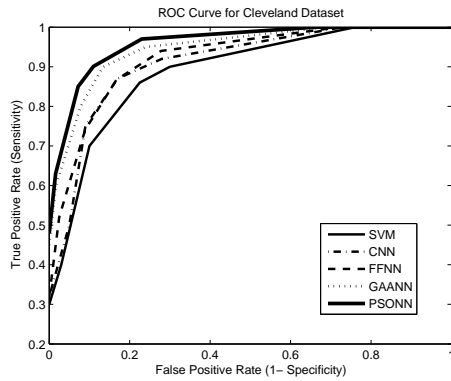


Fig. 7 ROC curve for Cleveland dataset.

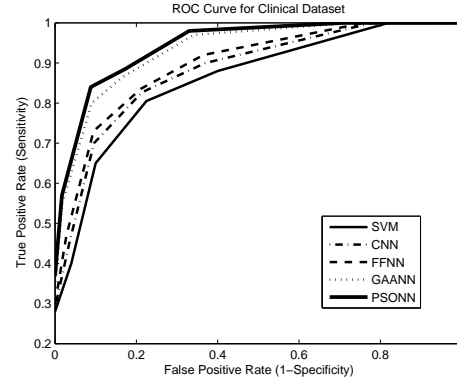


Fig. 8 ROC curve for Clinical dataset.

other classifier models. The result shows that the best area under the ROC curve is achieved for Cleveland database and real time clinical DB using PSONN. Therefore, PSONN have a great potential to be applied in automatic prediction of risk level of myocardial infarction by reducing the misclassification rate. Hence this model can be used as a diagnostic tool for cardiologist with high precision accuracy compared to other intelligent techniques described. In summary, it's found that the proposed PSONN based prediction system offers substantial improvement in prediction of heart disease. The results of the experiments show that an improvement of 4 % accuracy has been achieved by applying optimized neural network classifiers. This implies that the optimized neural network is one of the desirable classifier, which can be used as an aid for the physicians to predict the risk of MI.

7. Conclusions

This paper concentrates mainly in developing an intelligent computer aided detection system for Myocardial Infarction from patient record. The analysis of data is a challenging task due to the range and type of attribute selected. By adopting intelligent algorithms many classifiers are developed to characterize normal and abnormal patient from the given data. The dataset contains both normal and abnormal form patient record that is associated with ischemic heart disease. The selected attributes are compared by means of the ability to predict Myocardial Infarction using different intelligent algorithm. As a performance result, the PSONN classifier can be conveniently applied to detect MI, thus providing a second reader for the cardiologists. In future hybrid intelligent algorithm can be used to train the neural network for better classification.

Acknowledgement

The authors thankfully acknowledge the support and encouragement extended by N.I. Institute of Medical Science and Research Foundation and also for the necessary facilities provided.

References

- [1] BEGG R.K., PALANISWAMI M., OWEN B. Support vector machines for automated gait classification. *IEEE Trans. Biomed. Eng.* 2005, 52(5), pp. 828–838, doi: 10.1109/TBME.2005.845241.
- [2] CHAITRALI S.D., APTE S.S. Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques. *International Journal of Computer Applications*. 2012, 47(10), pp. 0975 – 888, doi: 10.5120/7228-0076.
- [3] CHITRA R., SEENIVASAGAM V. Knowledge Discovery from Heart Disease Dataset Using Optimized Neural Network. *Mining Intelligence and Knowledge Exploration*. 2013, vol. 8284, pp. 423–431, doi: 10.1007/978-3-319-03844-5_43.
- [4] DEEPIKA N., CHANDRA S.K. Association rule for classification of Heart Attack Patients. *International Journal of Advanced Engineering Science and Technologies*. 2011, 1(11), pp. 253–257.
- [5] DHEEBA J., PADMA A. Intelligent Adaptive Noise Cancellation using Cascaded Correlation Neural Networks. *International Conference on Signal Processing, Communications and Networking (ICSCN '07)*, Chennai, India. IEEE, 2007, pp. 178–182.
- [6] DING C.H.Q., DUBCHAK I. Multi-class protein fold recognition using support vector machines and neural networks. *Bioinformatics*. 2001, 17, pp. 349–358, doi: 10.1093/bioinformatics/17.4.349.
- [7] DYBOWSKI R., WELLER P., CHANG R., GANT V. Prediction of outcome in the critically ill using an artificial neural network synthesized by a genetic algorithm. *Oncology*. 1999, 52(4), pp. 281–286.
- [8] FAHLMAN S.E. Faster-Learning Variations on Back-Propagation: An Empirical Study. In: *Proceedings of the 1988 Connectionist Models summer School*. Morgan Kaufmann, 1988, pp. 38–51.
- [9] FIDELE B., CHEENEESHASH J., GOPAUL A., GOORAH S.S.D. Artificial neural network as a clinical decision-supporting tool to predict cardiovascular disease. *Trends in Applied Sciences Research*. 2009, 4(1), pp. 36–46, doi: 10.3923/tasr.2009.36.46.
- [10] FROST & SULLIVAN. Preparing for an aging society: Challengers faced by healthcare system in European Union, Japan and United States. San Antonio, TX, USA: Frost & Sullivan, 2009. Research report.
- [11] GUNN S.R. Support vector machines for classification and regression [online]. Southampton, U.K.: University of Southampton, 1998 [viewed 2015-12-30]. Technical report. Available from: <http://ce.sharif.ir/courses/85-86/2/ce725/resources/root/LECTURES/SVM.pdf>
- [12] HANDELS H., RO T.H., KREUSCH J., WOLFF H.H., POPPL S.J. Feature selection for optimized skin tumor recognition using genetic algorithms. *Artificial Intelligence in Medicine*. 1999, 16, pp. 283–297.
- [13] HECKERLING P.S., GERBER B.S., TAPE T.G., WIGTON R.S. Use of genetic algorithms for neural networks to predict community-acquired pneumonia. *Artificial Intelligence in Medicine*. 2004, 30, pp. 71–84, doi: 10.1016/S0933-3657(03)00065-4.
- [14] JABBAR M.A., CHANDRA P., DEEKSHATULU B.L. Cluster based association rule mining for heart attack prediction. *Journal of Theoretical and Applied Information Technology*. 2011, 32(2), pp. 196–201.
- [15] KAWAMOTO K., HOULIHAN C.A., BALAS E.A., LOBACH D.F. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ*. 2005, pp. 330–765, doi: 10.1136/bmj.38398.500764.8F.
- [16] KHANDOKER A.H., LAI D., BEGG R.K., PALANISWAMI M. Wavelet based feature extraction for support vector machines for screening balance impairments in the elderly. *IEEE Trans. Neural Syst. Rehabil. Eng.* 2007, 15(4), pp. 587–597, doi: 10.1109/TNSRE.2007.906961.
- [17] KONONENKO I. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine*. 2001, 23(1), pp. 89–109, doi: 10.1016/S0933-3657(01)00077-X.

- [18] KUOK K.K., HARUN S., SHAMSUDDIN S.M. Particle Swarm Optimization Feedforward Neural Network for Hourly Rainfall-runoff Modeling in Bedup Basin, Malaysia. *International Journal of Civil & Environmental Engineering*. 2009, 9(10), pp. 20–39, Available from: https://www.idc-online.com/technical_references/pdfs/civil_engineering/Particle Swarm.pdf
- [19] MONTANI S., BELLAZZI R., PORINALE L., STEFANELLI M. A multi-modal reasoning methodology for managing IDDM patients. *International Journal of Medical Informatics*. 2000, 58–59, pp. 243–256, doi: 10.1016/S1386-5056(00)00091-5.
- [20] MUNJAL Y.P., SHARMA S.K. *API textbook of medicine*. JP Medical Ltd, 2012.
- [21] LICHMAN M. UCI Machine Learning Repository [online]. Irvine, CA: University of California, School of Information and Computer Science, 2013 [viewed 2015-12-30]. Available from: <http://archive.ics.uci.edu/ml>
- [22] OZYILMAZ L., YILDIRIM T. Artificial neural networks for diagnosis of hepatitis disease. In: *Proceedings of the International Joint Conference on Neural Networks*. 2003, 1, pp. 586 – 589, doi: 10.1109/IJCNN.2003.1223422.
- [23] SAANGYONG U., KIM D.-H., KIM J., CHO S.W., CHEONG J.Y. Chronic Hepatitis Classification Using SNP Data and Data Mining Techniques. In: *Proceedings of the Frontiers in the Convergence of Bioscience and Information Technologies (FBIT 2007)*, Jeju City, South Korea. IEEE, 2007, pp. 81–86, doi: 10.1109/FBIT.2007.64.
- [24] SELLAP P., AWANG R. Intelligent Heart Disease Prediction System Using Data Mining Techniques. *IEEE/ACS International Conference on Computer Systems and Applications (AICCSA 2008)*, Doha, Qatar. IEEE, 2008, pp. 108–115, doi: 10.1109/AICCSA.2008.4493524.
- [25] SHANTHAKUMAR B.P., KUMARASWAMY Y.S. Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network. *European Journal of Scientific Research*. 2009, 31(4), pp. 642–656.
- [26] SRINIVAS K., RANI B.K., GOVRDHAN A. Application of Data Mining Techniques in Healthcare and Prediction of Heart Attacks. *International Journal on Computer Science and Engineering*. 2011, 2(2), pp. 250–255. Available from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.163.4924&rep=rep1&type=pdf>
- [27] TAREK H., AL-HARTHI M.M., M.T. FAHEEM. Adaptive Ensemble and Hybrid Models for Classification of Bioinformatics datasets. *Global Journal of Technology and Optimization*. 2012, 3, pp. 20–29.
- [28] WORLD HEALTH ORGANIZATION. Future Trends In Global Mortality: Major Shifts In Cause Of Death Patterns [online]. In: *World Health Statistics*. Geneva: WHO Press, 2008, pp. 29–31. Available from: http://www.who.int/entity/gho/publications/world_health_statistics/EN_WHS08_Full.pdf
- [29] XING Y., WANG J., ZHAO Z., GAO Y. Combination data mining models with new medical data to predict outcome of coronary heart disease. In: *Proceeding of International Conference on Convergence Information Technology*, Gyeongju, South Korea. IEEE, 2007, pp. 868–872, doi: 10.1109/ICCIT.2007.204.