Analysis and Evaluation of Techniques for Myocardial Infarction Based on Genetic Algorithm and Weight by SVM

Hodjatollah Hamidi*
Department of Industrial Engineering, K. N. Toosi University of Technology, Tehran, Iran
h_hamidi@kntu.ac.ir
Atefeh Daraei
Department of Industrial Engineering, K. N. Toosi University of Technology, Tehran, Iran
adaraei@mail.kntu.ac.ir

Abstract

Although decreasing rate of death in developed countries because of Myocardial Infarction, it is turned to the leading cause of death in developing countries. Data mining approaches can be utilized to predict occurrence of Myocardial Infarction. Because of the side effects of using Angioplasty as main method for diagnosing Myocardial Infarction, presenting a method for diagnosing MI before occurrence seems really important. This study aim to investigate prediction models for Myocardial Infarction, by applying a feature selection model based on Wight by SVM and genetic algorithm. In our proposed method, for improving the performance of classification algorithm, a hybrid feature selection method is applied. At first stage of this method, the features are selected based on their weights, using weight by SVM. At second stage, the selected features, are given to genetic algorithm for final selection. After selecting appropriate features, eight classification methods, include Sequential Minimal Optimization, REPTree, Multi-layer Perceptron, Random Forest, K-Nearest Neighbors and Bayesian Network, are applied to predict occurrence of Myocardial Infarction. Finally, the best accuracy of applied classification algorithms, have achieved by Multi-layer Perceptron and Sequential Minimal Optimization.

Keywords: Artificial Neural Network; Sequential Minimal Optimization; REPTree; Knowledge Discovery in Databases; Myocardial Infarction.

1. Introduction

Myocardial Infarction (MI) is a prominent cause of death all around the world [1]. In spite of reduction in mortality rates because of cardiovascular diseases in developed countries, it is known as a significant cause of death in developing countries [2]. Based on [3] it is likely to ischemic heart disease become the most common cause of death by 2020 [3]. Myocardial Infarction is turned to the main cause of death for Iranian people above 35 years [4]. As mentioned in [5], Ischemic heart disease is known as the main cause of death in Western countries. Ischemic heart disease is happened due to Atherosclerosis. Atherosclerosis is the process of depositing Cholesterol and fat in the vessels especially in people with genetic susceptibility, overweight and obesity, inactive way of life, high blood pressure, and people who consume a lot of cholesterol and fat. This is lead to production of atherosclerotic plaques which cause the fully or partially blockage of blood [5]. Based on Ministry of Health and Medical Education report in 2012, the rate of mortality because of Myocardial Infarction is estimated 85 per 100,000 [2]. Myocardial Infarction means the death of heart muscle due to a sudden blockage of a coronary artery by a blood clot [6]. Thus, immediately after coronary occlusion, blood flow is stopped in the coronary vessels beyond the blocked site with the exception of a very small amount of collateral blood flow in surrounding vessels. So an area of the heart muscle with zero blood flow or with very low blood flow to do muscle action has been infarcted. This process is called Myocardial Infarction [5].

Data mining methods transform raw data into useful information [7]-[8]. Therefore, data mining can be regarded as a tool for acquiring knowledge from raw and meaningless data in the medical field, and in the field of Myocardial Infarction in particular. Data mining is a part of knowledge discovery in databases (KDD) [9]; in [10] data mining is introduced as a KDD process includes three stages: Data Preprocessing, Data Modeling and Data Post Processing [10]. Data modeling tasks in the data mining process are divided into two categories: predictive tasks and descriptive tasks. The algorithms of the predictive category include classification algorithms for discrete data, and regression algorithms, for continuous data [11], which are learned through supervised learning process [7].

The algorithms which determine the class labels based on the training data with particular labels are called classification algorithms [7]. Early detection of occurrence Myocardial Infarction seems to be very important because of its important role of mortality in the world and Iran. Angiography is a costly and non-invasive method for detecting the blockage in vessels and may have complications for patients such as bleeding at the insertion

site to the vein, stroke, vascular injury, renal failure and even death. Thus, due to side effects of drugs and angiography [12], using prediction methods could consider really important. The aim of this study is early prediction of Myocardial Infarction using the classification algorithm and a proposed feature selection. In addition, a combined feature selection method was used along with the classification techniques. To the best of our knowledge, using genetic algorithm along with feature selection method has not been used so far. The algorithms Naïve Bayes, Support Vector Machine, Artificial Neural Network, K-Nearest Neighbors, Sequential Minimal Optimization, REPTree, Random Forest and Bayesian Network are used for reaching to this purpose. The paper is organized as follows: In second section the related researches in this field are reviewed. Section three describe the Myocardial Infarction data set. Section four explain the proposed model. Experimental Results comes in fifth section. In the section sixth the results are analyzed. The section seventh contains conclusion.

2. Literature Review

Many researches have been worked in this field. In a study, Baxt et al. in 2002, proposed a model for prediction of acute Myocardial Infarction. They used artificial Neural Network method. Data set used in this study consisted of 2204 patient with 40 attributes. The results showed that the Neural Network has a high potential for use in prediction of acute Myocardial Infarction [13]. The authors in [14] presented a system which detects abnormality in the heart and its walls. They used a real data set which contains 141 samples. After Feature selection, Support Vector Machine (SVM) algorithm was applied to the data. The results show that the selection of 3 main features selected by feature selection leads to high efficiency.

Conforti et al. in [15] employed a Support Vector Machine with 5 core functions include linear, Gaussian, Laplacian, polynomial and sigmoid. The aim was to categorize the patterns for detection of acute Myocardial Infarction (AMI). They used a dataset consisted of 242 patients with chest pain. 105 features in this data set were divided to 4 categories: history, ECG, electrocardiogram, and blood test. The best accuracy in 7 different Feature selection modes, on average, was equal to 85.85% for linear and polynomial function, 82.64% for the Gaussian, 86.43% for Laplace and 82.4% for the sigmoid. In [16], is used a Neural Network for diagnosing heart attacks. UCI datasets is used in this paper, consisted of 13 features. After preprocessing, clustering is used to determine categories. After the rules extracted in clusters using Mafia algorithm, a three layered Neural Network is applied on data. In [17], a model is presented to predict Myocardial Infarctions and coronary artery bypass graft. The algorithm C4.5 decision tree is used for classification. The data set consisted of 1200 patients. Size of the data set is reduced to 369 after excluding the inappropriate

samples. After applying C4.5 tree, they extracted the relevant rules and the most important factors affecting the occurrence of Myocardial Infarction. In 2010, Arif et al. used Back Propagation Neural Network for detection Myocardial Infarction MI and determine its location. The main features was Q and T waves as well as ST-Elevation or ST-Depression. The data set used in this study consisted of 148 MI patients. Neural Network algorithm was used for both detection MI and determining its location. Sensitivity and specificity measures were used to assess the results of detection of MI (97.5% and 99.1%, respectively) [18]. In [19], 5 classification algorithms is employed to develop a model for prediction of heart attack The algorithms used for this purpose included J48 decision tree, Bayesian Network, Naïve Bayes(NB), CART and REPTREE. 11 features were investigated in this study. According to the experimental results, J48, Naïve Bayes and CART classification methods achieved higher accuracy (99.07%) and proved to be better than Naïve Bayes techniques and Bayesian Network. In 2014, the authors used a model of multi-layered feed-forward Neural Network algorithm predicting heart attack, which utilized genetic algorithm for initialization and discovering the optimal weights for multi-layered feed forward Neural Network. The data set used in this study, obtained from UCI, which includes 270 cases and 13 features. Finally, in the evaluation stage, the propose method obtained accuracy of 88% [20]. In [21] is attempted to detect and locate MI using 4 algorithms include Probabilistic Neural Network (PNN), K-Nearest Neighbors (K-NN), Multi-layer Perceptron (MLP), and Naïve Bayes. The dataset used in this study included information of 549 patients with 15 features. After applying the algorithms to the data the highest accuracy for the detection of MI was obtained by the Naïve Bayes method, which was equal 94.74%. The obtained accuracy was lower for the diagnosis of MI. The author in [22] aimed to classify ECG signals as healthy or Myocardial Infarction. The data set used in this study was the PTB data set, included 82 healthy cases and 367 cases of Myocardial Infarction. After applying two methods of artificial Neural Network and Support Vector Machine on the data set, the results showed that SVM has higher accuracy than the artificial Neural Network. Sharma et al. in 2015, proposed a model for diagnosing Myocardial Infarction using K-Nearest Neighbors and SVM. The kernel functions Lin and RBF used for SVM algorithm. The accuracy obtained for K-NN was equal 81%. The accuracy resulted in using SVM was higher, so that SVM with Lin function obtained the accuracy 89% and SVM with RBF achieved to highest accuracy equal to 96% [23]. Kora and Kalva in 2015, proposed a model of Neural Networks (LN and SCG), K-NN and SVM combined with improved BAT algorithm for diagnosing heart attack. They used ECG features of PTB data base. This data base, used in this study, consisted of 52 normal cases and 148 MI. BAT algorithm was used for feature extraction. The results showed that, the highest accuracy

belongs to Neural Network (LM) which is equal to 98.1%. The other algorithms' accuracy consisted of 87.9% for Neural Network (SCG), 65.1% for K-NN and 76.74% for SVM [24].

3. Myocardial Infarction Data Set Description

In this study, a dataset consisted of the information collected from 519 visitors to Shahid Madani Specialized Hospital of Khorram Abad, include 222 without MI and 297 with MI, is used. After reviewing patient records and consulting with specialists, and based on [3] and [12] books, features such as, Troponin I, CRP, LDH and three main heart arteries include LAD, RCA and LCX, were added to the features in articles reviewed. The dataset contains 52 features (51 predictor features and 1 goal feature), presented in table 1.

4. The Proposed Model

4.1 Preprocessing Process

Preprocessing method is an important part of data mining process which prepare the raw data for data mining [25]. Raw data is usually incomplete, noisy and incompatible [26] because of huge size of databases, integration of several various datasets or human errors [27]. In this study the missing values are filled using the same available values. Moreover, normalization, as a method for converting the data into a suitable form for data mining, is used in this study. In this study, the Min - Max normalization method on the interval of [0, 1] is used.

Table 1. The Features of Myocardial Infarction Data Set

Features	Range
Age	28-93
Sex	M, F
Weight	43-120
Body Mass Index (BMI)	16-42
Systolic Blood Pressure (SBP)	$80 - 210$
Diastolic Blood Pressure (DBP)	$40 - 190$
Heart Rate	$50 - 190$
Family History	Yes , No
Smoke	Yes , No
Obesity	Yes , No
Hypertension	Yes , No
Chronic Renal Failure (CRF)	Yes , No
Cerebrovascular Accident (CVA)	Yes , No
Congestive Heart Failure (CHF)	Yes , No
Dyslipidemia (DLP)	Yes , No
Blood Pressure	Yes , No
Edema	Yes , No
Fatigue and weakness	Yes , No
Lung Rales	Yes , No
Typical Chest Pain	Yes , No
Distribution of pain to arms and neck	Yes , No
Dyspnea	Yes , No
Atypical CP (Atypical Chest Pain)	Yes , No
Non-anginal CP (Non-anginal Chest Pain)	Yes , No
Exertional CP (Exertional Chest Pain)	Yes , No

Features	Range
Troponin I	Pos.–Neg.
C-reactive protein (CRP)	Pos. – Neg.
Total Cholesterol	$22 - 480$
Fasting Blood Sugar (FBS)	$23 - 550$
Lactate Dehydrogenase (LDH)	$30 - 1000$
Creatine Phosphokinase (CPK)	$8 - 650$
Creatinine (Cr)	$0.3 - 18$
Triglyceride (TG)	$70 - 400$
Low Density Lipoprotein (LDL)	$38 - 182$
High Density Lipoprotein (HDL)	$21 - 102$
Ejection Fraction (EF)	$10 - 80$
Lymphocyte	2- 81
Platelet	$90 - 1700$
Blood Urea Nitrogen (BUN)	$10 - 80$
Erythrocyte Sedimentation Rate (ESR)	$4 - 76$
Hemoglobin	$3.5 - 23$
Potassium (K)	$2.5 - 12.5$
Sodium (Na)	$62 - 192$
White Blood Cells (WBC)	3200- 20000
ST Elevation	Yes , No
ST Depression	Yes , No
T inversion	Yes , No
Poor R Progression	Yes , No
LAD (left Anterior Descending artery)	Yes , No
LCX (Left Coronary Artery)	Yes , No
RCA (Right Coronary Artery)	Yes , No

4.2 Proposed Feature Selection Method

Feature selection is one of the most common used methods of dimensionality reduction. In this methods after removing the irrelevant features, the best features is kept and used [9].

At the first stage of the proposed feature selection method used in this study, weight by SVM is applied. Using this operator, first the features' weights are specified. These features, then, are used in Genetic Algorithm, at the second stage. Genetic algorithm finds the best hypothesis by searching a hypothesis space. The initial hypotheses called a population is randomly generated and a fitness function is used for evaluating each hypothesis. Hypotheses with greater fitness have the higher probability of being chosen to create the next generation. Some of the best hypotheses may be retrained at the next generation, the other operations, crossover and mutation are used to generate new hypotheses. The size of population is same for all generations [28]. To the best of our knowledge, in subject of Myocardial Infarction prediction, the two using Genetic Algorithm along with feature selection is not used so far.

4.3 Classification Algorithms

Classification algorithms employed in this study, to assess the performance of proposed feature selection, consist of REPTree, Random Forest, Bayesian Network, Support Vector Machine, Multi-layer Perceptron, K-Nearest Neighbors, Sequential Minimal Optimization and Naïve Bayes.

4.3.1 RRPTree

REPTree utilize the regression tree logic and creates many trees in various iterations. Next it chooses the best

tree from all created trees. In general, Reduced Error Pruning Tree (REPTree) is a fast decision tree learning method that create the trees in iterations and prunes them using reduced error pruning [29].

4.3.2 Random Forest

Random forest is a classifier contain of a combination of trees. Each tree is produced of a random vector sampled of the input vector. Every tree considered as a unit, which vote for the most popular class to classify instance [30].

4.3.3 Bayesian Network

Bayesian Network is a network based on the relationship between attributes. It utilizes statistic techniques to represent probability this relationships. This algorithm, similar to Naïve base, uses Bayes rule [31].

4.3.4 Multi-Layer Perceptron (MLP)

Neural Network is capable of predicting new observations, from earlier ones, after executing the learning process using the past data [32]. One of the mostly used learning algorithm for Neural Networks is Multilayer Perceptron [33].

4.3.5 K-Nearest Neighbors (K-NN)

K-Nearest Neighbors algorithm is based on comparing a given test instance with training instances which are similar to it. When an unknown instance is given, the algorithm searches the pattern space for the k training instances that are closest to the unknown instance. These k training instances are the k "nearest Neighbors" of the unknown instance [26].

4.3.6 Support Vector Machine (SVM)

SVM is a classification algorithm, which is based on the statistical learning theory [34]. This algorithm maps input vectors to higher dimensional spaces where maximal separating hyper-planes are constructed. Two parallel hyper-planes are constructed on both sides of a data searching hyper-plane. The separating hyper-plane is the one that maximizes distance between two parallel hyper-planes [35].

4.3.7 Sequential Minimal Optimization (SMO)

SMO is used for learning SVM algorithm. This algorithm decrease the time of obtaining the weights for SVM, in optimization problems. This low time is due to using the serial optimizing methods and the linear memory that this methods require. SMO doesn't need ant matrix storage. Since no matrix algorithms are employed in SMO, its sensitivity to numerical problems' precision is low [36].

4.3.8 Naïve Bayes (NB)

The Naïve Bayes algorithm is a probabilistic classifier based on conditional probability. It means, the Naive Bayes classifier uses probability to classify the new instance [9]. The algorithm uses of all the features contained in the dataset, and based on its main assumption, Conditional

independence. In fact, it considers the features equally important and independent of each other [37].

5. Experimental Results

In this study, version 7.0.1 of RapidMiner is used for classification. To evaluate the performance of the classification models accuracy, sensitivity and specificity can be used, which are obtained using K-fold cross validation. They are the most popular measures for evaluating the classification performance.

Classification accuracy is one of the most common metrics for evaluating performance of the model. It is the ratio of TP and TN obtained by model to the total number of instances [38], as shown by equation (3):

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \qquad (1)$$

where TN, TP, FP and FN represents the number of true negatives, true positives, false positives and false negatives, respectively [39].

Sensitivity and specificity are also two performance measures for evaluating of the efficiency of a classification technique. Sensitivity, or true positive rate, is the ratio of positive instances have been truly classified as positive. Specificity is the ratio of negative instances have been truly classified as negative [40].

5.1 Results

This section present the experimental results of the proposed classification models. These algorithms are compared to each other in 2 different statues, using or not using of feature selection method. The input parameter for selection of features using weight by SVM is set to "top p%", where p value is equal to 0.7. It means the top 70% of all features are selected.

For genetic algorithm the best results are obtained when the number of generations is set to 10 and the population size set to 7. The other parameters are used at the software default. After using Genetic Algorithm, it selects 32 features from the 36 features selected using weight by SVM. The weights of the features selected in first stage and selected features after implementing second stage of feature selection is shown in Table 2. Third column of this table represent the selected features at end of the feature selection model.

Thus, the final features selected by proposed feature selection method include: CHF, Troponin I, CRP, Total cholesterol, LDL, HDL, ESR, Lymphocyte, BMI, sex, Family History, Obesity, CRF, CVA, FBS, Cr, TG, BUN, Hemoglobin, Na, EF, Edema, Lung rales, Typical chest pain, Dyspnea, Non-anginal CP, Exertional CP, ST Depression, T inversion, ST Elevation, Poor R Progression, LAD.

It is worth noting that the parameter K is set to 4. The other algorithms are implemented in defaults of software. Table 3 shows the results for implemented classification algorithms on

Myocardial Infarction dataset, in two states: using proposed feature selection method and not using feature election.

Table 2. Selected Features by Proposed Feature Selection Method

Attribute	Weights of selected features after first stage of FS	Selection status after second stage of FS
ST Elevation	1.0	✓
Troponin I	0.98	✓
ST Depression	0.96	✓
CRF	0.94	✓
Exertional CP	0.92	✓
T inversion	0.9	✓
Obesity	0.88	✓
K	0.86	×
TG	0.84	✓
HDL	0.82	✓
ESR	0.8	✓
FBS	0.78	✓
Lymphocyte	0.76	✓
Edema	0.74	✓
Hemoglobin	0.72	✓
Lung rales	0.7	✓
BUN	0.68	✓
EF	0.66	✓
Dyspnea	0.64	✓
Poor R Progression	0.62	✓
LDL	0.6	✓
Non-anginal CP	0.58	✓
Typical chest pain	0.56	✓
Atypical CP	0.54	×
heart rate	0.52	×
Na	0.5	✓
Family History	0.48	✓
CVA	0.46	✓
CHF	0.44	✓
sex	0.42	✓
Total cholesterol	0.4	✓
DLP	0.38	×
LAD	0.36	✓
CRP	0.34	✓
BMI	0.32	✓
Cr	0.3	✓

Table 3. Performance of Classification Algorithms

Algorithm		Accuracy	Sensitivity	Specificity
NB	not using FS	95.95	95.96	95.95
	using FS	96.72	96.97	96.40
MLP	not using FS	96.33	95.96	97.75
	using FS	97.50	97.31	97.52
SVM	not using FS	96.15	95.96	96.40
	using FS	95.76	96.30	95.50
K-NN	not using FS	95.75	93.27	99.10
	using FS	97.69	96.30	99.55
SMO	not using FS	96.15	96.30	95.95
	using FS	97.30	97.31	97.30
REPTree	not using FS	94.61	93.94	95.50
	using FS	95.38	93.94	97.30
Random Forest	not using FS	90.37	97.98	80.18
	using FS	93.45	97.64	87.84
Bayesian Network	not using FS	95.57	95.29	95.95
	using FS	96.34	95.62	97.30

6. Analysis of the Results

As shown in Table 3, K-NN, MLP and SMO algorithms have the best accuracy. Besides, although the accuracy of these algorithms are the best, the other algorithms have achieved the accuracy above 93%. The chart in Figure 1 provide a comparison of accuracy of algorithms in two state using or not using of feature selection. From this chart it is obvious that using proposed feature selection method improved the accuracy of all the algorithms, except SVM. It is noting that although RF algorithm have achieved the lowest accuracy in comparison with the other algorithms, the 3% improvement of the accuracy of this algorithm after using feature selection method is notable.

The results also show that using the feature selection method have enhanced the sensitivity of algorithms, which means the tendency of algorithms to classify MI cases is increased. In terms of specificity measure, except for SVM, feature selection have led to higher results; it means that the trend of algorithms to predict heathy cases in increased. It is worth mentioning that using the feature selection method have caused a much improvement in specificity for RF algorithm, which is about 9%.

The results achieved in this study in comparison to the studied, reviewed in section 2, Show the better performance of proposed method. In compared to [14 -15] and [22 - 24], which have used SVM, despite of reduction of accuracy after using feature selection, our implemented model with SVM algorithm achieved higher accuracy.
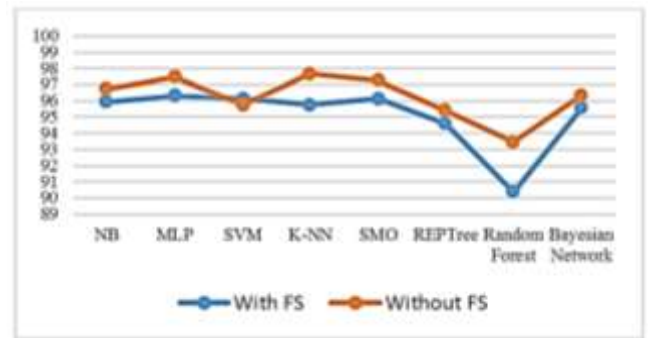

Fig. 1. Comparison of classification algorithms accuracy in two state: using or not using proposed feature selection

Moreover, our model using K-NN in compared to [21] and [23-24] have reached better accuracy, so that the accuracy of our model using K-NN is about 9% higher than the reviewed papers. The model in this study using MLP have reached better accuracy in comparison with the results of the models proposed in [13], [20 -22], so that the accuracy of our model using MLP is about 9% higher than the best accuracy which is resulted in [21].

Although Naïve Bayes algorithm's result in this study is higher than the accuracy in [21], it is lower than the accuracy resulted in [19]. Also the accuracy of Bayesian Network and REPTree in our study is lower than the accuracies achieved in [19]. This being low of accuracy in some algorithms in our study may be due to the difference between the number and types of features and cases used in studies which can affect the results.

7. Conclusion

In this study a new feature selection method is proposed which utilized Weight by SVM and Genetic Algorithm. Eight classification algorithms, include Naïve Bayes, Multi-layer perceptron, Support Vector Machine, K-Nearest Neighbors, Bayesian Network, SOM, Random forest and REPTree, are applied to a real Myocardial Infarction. Findings showed that using this feature selection method can lead to higher accuracy for classification algorithms in predicting Myocardial Infarction. After applying the proposed feature selection method, a subset of the features selected and the results showed increasing in accuracy, sensitivity and specificity in all algorithms, except SVM. Among the investigated algorithms, K-NN (K=4), MLP and SOM algorithms, had the best accuracies, 97.69%, 97.50% and 97.30% respectively. Overall, it can be concluded that applying the feature selection method, used in this study, along with classification algorithms is almost a confident method for predicting Myocardial Infarction. High sensitivity and specificity, in addition to high accuracy of the algorithms, can be considered as a benefit of the proposed model. One weakness can be considered about work may be the large number of MI cases in compared to normal cases; because, in general, disease datasets are usually imbalanced and the number of disease cases are much less than normal cases. This problem could be effective on the results. In the future we are going to employ more classification algorithms using the proposed feature selection method by considering more normal cases than healthy cases. Besides, we want to make changes in our feature selection method using the other evolutionary method and weighting methods and compared the results. As a suggestion this method can be used for the other health problems, such as diseases prediction.

References

[1] A.S. Go, D. Mozaffarian, V.L. Roger, E.J. Benjamin, J.D. Berry, W.B. Borden, D.M. Bravata, S. Dai, E.S. Ford, C.S. Fox, and S. Franco."Heart Disease and Stroke Statistics--2013 Update: A Report From the American Heart Association". Circulation, Vol. 127, pp. e6-e245, 2012.

[2] A. Ahmadi, H. Soori, Y. Mehrabi, K. Etemad, T. Samavat, and A. Khaledifar."Incidence Of Acute Myocardial Infarction In Islamic Republic Of Iran: A Study Using National Registry Data In 2012". Eastern Mediterranean health journal, Vol. 21, pp. 5-12, 2015.

[3] C. Wiener, C. Brown, A. Hemnes and T. Harrison. Harrison's principles of internal medicine. New York: McGraw-Hill Medical, 2012, pp. 455-456.

[4] F. Mohammadi, A. Taherian, M. Hosseini and M. Rahgozar. "Effect of Home-Based Cardiac Rehabilitation on Quality of Life in the Patient with Myocardial Infarction". Journal of Rehabilitation, Vol. 7, pp. 11-19, 2006. [In Persian]

[5] J. E. Hall. Guyton and Hall Textbook of Medical Physiology. New York: Saunders, 2015, pp. 264-266.

[6] R. Dhingra, J. Shaw and L. A. Kirshenbaum. "molecular regulation of apoptosis signaling pathway in heart" in Apoptosis: Modern Insights into Disease from Molecules to Man, 1st ed., V. R. Preedy, Ed. Florida: CRC Press, 2010, pp. 382-385.

[7] N. Esfandiari, M. Babavalian, A. Moghadam and V. Tabar. "Knowledge discovery in medicine: Current issue and future trend". Expert Systems with Applications, Vol. 41, pp. 4434-4463, 2014.

[8] M. Jabbar, B. Deekshatulu and P. Chandra. "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm". Procedia Technology, Vol. 10, pp. 85-94, 2013.

[9] S. Kumar and G. Sahoo. "Classification of Heart Disease Using Naïve Bayes and Genetic Algorithm". Computational Intelligence in Data Mining, Vol. 2, pp. 269-282, 2014.

[10] U. Fayyad and R. Uthurusamy. "Data mining and knowledge discovery in databases". Communications of the ACM, Vol. 39, pp. 24-26, 1996.

[11] P. Tan, M. Steinbach and V. Kumar. Introduction to data mining. Boston: Pearson Addison Wesley, 2005.

[12] I. Benjamin, R. C. Griggs, E. J Wing and J. Fitz. Andreoli and Carpenter's Cecil Essentials of Medicine.New York: Saunders, 2015, pp. 93-102.

[13] W. Baxt, F. Shofer, F. Sites and J. Hollander. "A neural computational aid to the diagnosis of acute Myocardial Infarction". Annals of Emergency Medicine, Vol. 39, pp. 366-373, 2002.

[14] M. Qazi, G. Fung, S. Krishnan, J. Bi, R. Bharat Rao and A.S. Katz. "Automated heart abnormality detection using sparse linear classifiers". Engineering in Medicine and Biology Magazine, Vol. 26, pp. 56-63, 2007.

[15] D. Conforti, D. Constanzo and R. Guido. "Medical decision making: A case study within the cardiology domain". Journal on Information Technology in Healthcare, Vol. 5, pp. 343–356, 2007.

[16] S. Patil and Y. Kumaraswamy. "Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network". European Journal of Scientific Research, Vol. 31, pp. 642-656, 2009.

[17] M. Karaolis, J. Moutiris, L. Papaconstantinou and C. Pattichis. "Association rule analysis for the assessment of the risk of coronary heart events." in Engineering in Medicine and Biology Society. EMBC 2009. Annual International Conference of the IEEE, 2009, pp. 6238 - 6241.

[18] M. Arif, I. Malagore and F. Afsar. "Automatic Detection and Localization of Myocardial Infarction Using Back Propagation Neural Networks", in 4th International Conference on Bioinformatics and Biomedical Engineering (iCBBE), 2010, pp. 2151-7614.

[19] H. Masethe and M. Masethe, "Prediction of Heart Disease using Classification Algorithm," in Proceedings of the World Congress on Engineering and Computer Science, San Francisco, USA, 2014.

[20] N. Krishnaraj and R. Vinothkumar. "Heart Disease Prediction using GA and MLBPN". International Journal of Applied Management & Business Utility, Vol. 2, pp. 17-24, 2014.

[21] N. Safdarian, N. Dabanloo and G. Attarodi. "A new pattern recognition method for detection and localization of Myocardial Infarction using t-wave integral and total integral as extracted features from one cycle of ECG signal". Journal of Biomedical Science and Engineering, Vol. 07, pp. 818-824, 2014.

[22] N. Bhaskar. "Performance analysis of Support Vector Machine and Neural Networks in detection of Myocardial Infarction". Procedia Computer Science, Vol. 46, pp. 20-30, 2015.

[23] L. Sharma, R. Tripathy and S. Dandapat. "Multiscale Energy and Eigenspace Approach to Detection and Localization of Myocardial Infarction". IEEE Transactions on Biomedical Engineering, Vol. 62, pp. 1827-1837, 2015.

[24] P. Kora and S. Kalva. "Improved Bat algorithm for the detection of Myocardial Infarction". SpringerPlus, Vol. 4, pp. 1-18, 2015.

[25] U. Fayyad and R. Uthurusamy. "Data mining and knowledge discovery in databases". Communications of the ACM, Vol. 39, pp. 24-26, 1996.

[26] N. G. B. Amma. "Cardiovascular disease prediction system using genetic algorithm and Neural Network," in International Conference on Computing, Communication and Applications (ICCCA), 2012, pp. 1-5.

[27] Han J, Kamber M, Pei J. Data mining: concepts and techniques. Morgan Kaufmann, 2011.

[28] M. Pacharne and V. Nayak. "Feature Selection Using Various Hybrid Algorithms for Speech Recognition," in Computational Intelligence and Information Technology, 1st ed., V. Das and N. Thankachan, Ed. Berlin: Springer Berlin Heidelberg, 2011, pp. 652-656.

[29] S. Kalmegh. "Analysis of WEKA Data Mining Algorithm REPTree, Simple Cart and RandomTree for Classification of Indian News". IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2, pp. 438-446, 2015.

[30] M. Pal. "Random forest classifier for remote sensing classification". International Journal of Remote Sensing, Vol. 26, pp. 217-222, 2005.

[31] Y. Wang and J. Vassileva, "Bayesian Network-based trust model." in International Conference on Web Intelligence, IEEE/WIC, 2003, pp.372-378.

[32] R. Ganesh Kumar and Y. Kumaraswamy. "Performance Analysis Of Soft Computing Techniques For Classifying Cardiac Arrhythmia". Indian Journal of Computer Science and Engineering (IJCSE), Vol. 4, pp. 459-465, 2014.

[33] M. Gardner and S. Dorling. "Artificial Neural Networks (the multilayer perceptron)—a review of applications in the atmospheric sciences". Atmospheric Environment, Vol. 32, pp. 2627-2636, 1998.

[34] K. Polat, S. Güneş and A. Arslan. "A cascade learning system for classification of diabetes disease: Generalized Discriminant Analysis and Least Square Support Vector Machine". Expert Systems with Applications, Vol. 34, pp. 482-487, 2008.

[35] S. Gunn. "Support Vector Machines for Classification and Regression". University of Southampton, Technical Report, 1998.

[36] J. Platt. "Sequential minimal optimization:A fast algorithm for training Support Vector Machines". Microsoft Research, Technical report MSR-TR-98-141998.

[37] D. Hand and K. Yu. "Idiot's Bayes: Not So Stupid after All?", International Statistical Review/Revue Internationale de Statistique, Vol. 69, p. 385, 2001.

[38] R. Alizadehsani, J. Habibi, B. Bahadorian, H. Mashayekhi, A. Ghandeharioun, R. Boghrati and Z. Alizadeh Sani. "Diagnosis of Coronary Arteries Stenosis Using Data Mining", Journal of Medical Signals and Sensors, Vol. 2, pp. 153-159, 2012.

[39] A. Onan. "A fuzzy-rough nearest neighbor classifier combined with consistency-based subset evaluation and instance selection for automated diagnosis of breast cancer". Expert Systems with Applications, Vol. 42, pp. 6844-6852, 2015.

[40] M. Heydari, M. Teimouri, Z. Heshmati and S. Alavinia, "Comparison of various classification algorithms in the diagnosis of type 2 diabetes in Iran", International Journal of Diabetes in Developing Countries, 2015, pp.1-7.

Hodjatollah Hamidi born 1978, in shazand Arak, Iran, He got his Ph.D in Computer Engineering. His main research interest areas are Information Technology, Fault-Tolerant systems (fault-tolerant computing, error control in digital designs) and applications and reliable and secure distributed systems and E- Commerce. Since 2013 he has been a faculty member at the IT group of K. N. Toosi University of Technology, Tehran Iran. Information Technology Engineering Group, Department of Industrial Engineering, K. N. Toosi University of Technology.

Atefeh Daraei born in Khorram Abad, Lorestan Iran. She received her B.Sc in Information Technology in 2011 from University College of Nabi Akram, Tabriz, Iran. She is currently an M.Sc student in Information Technology (E-Commerce), at K. N. Toosi University of Technology, Tehran Iran. Her research interests include Machine learning, Knowledge Discovery and Data Mining and Customer Relationship Management.