Predicting Heart Attacks in Patients Using Artificial Intelligence Methods

Abbas Nasrabadi[1] & Javad Haddadnia[2]

[1] Mechatronics Engineering Department, Computer and Electrical Engineering Faculty, Hakim Sabzevari University, Sabzevar, Iran

[2] Associate Professor of Medical Engineering Department, Computer and Electrical Engineering Faculty, Hakim Sabzevari University, Sabzevar, Iran

Correspondence: Abbas Nasrabadi, Mechatronics Engineering Department, Computer and Electrical Engineering Faculty, Hakim Sabzevari University, Sabzevar, Iran.

Abstract

Today the heart disease is one of the most important causes of death in the world. So its early prediction and diagnosis is important in medical field, which could help in on time treatment, decreasing health costs and decreasing death caused by it. In fact the main goal of using data mining algorithms in medicine by using patients' data is better utilizing the database and discovering tacit knowledge to help doctors in better decision making.

Therefore using data mining and discovering knowledge in cardiovascular centers could create a valuable knowledge, which improves the quality of service provided by managers, and could be used by doctors to predict the future behavior of heart diseases using past records. Also some of the most important applications of data mining and knowledge discovery in heart patients system includes: diagnosing heart attack from various signs and properties, evaluating the risk factors which increases the heart attack.

In this article the effort focused on evaluating the previous works on discovering knowledge using data mining in heart diseases field, and also explain the used algorithms in every one of the previous works, to help the future researchers to gain maximum benefits from these abilities. Because of this, in the next sections, first we will explain various works in data mining field using heart patients' data, and will show the ability of data mining in various applications of heart disease field, and based on a table will show the history of data mining and it's applications in heart diseases field. Finally we will provide the best methods and algorithms used in various applications of heart diseases using a comparison and will show the results in a table. It is obvious in the diagrams that the suggested method has the best performance and best quality in prediction.

Keywords: heart diseases, fuzzy clustering, multi-objective genetic algorithm, silhouette index, dunn factor

1. Introduction

Daily increasing development in information technology caused in significant growth in sciences. One of the sciences is medical science. Using artificial intelligence techniques in all subjects of this branch of science especially cardiovascular diseases made it possible to design medical assistant systems. By taking attention to increase in new diseases and also extension of technologies, the diagnosis of diseases gone beyond the internal treatment style, and the most efforts of doctors and specialists is focused on early prediction of diseases using available signs. Medical information retrieval system is the best system for managing clinical data. This system is capable to healthcare operations in diagnosing diseases and has an important role in clinical decision making [1].

Cardiovascular diseases is one of the most spreading causes of death in worldwide. One main type of this disease is "coronary artery disease" (CAD), which about 25% of population without any previous signs, are suddenly subject of this disease, and experience severe heart attack and die [2]. At the moment, Angiography uses for determining the amount and location of narrowing of the arteries of the heart, which has high price and several side effects. Using data mining for diagnosis of heart diseases may be very lower in price and very faster [3].

Based on the announced statistics by the World Health Organization in 2005, there was 17.5 million victims from

cardiovascular diseases, which is 30% of all death in worldwide, and it was predicted that this value increase to 23 million people up to 2030. Examinations made in Iran showed that 38% of all death subjects caused by cardiovascular diseases, which is increasing in future. Based on the statistics obtained from the evaluation of cardiovascular diseases, it was shown that 16.1% of people have high blood pressure, 43.9% have extra weight, 38.9% have low physical activity, and also 10.8% use tobacco which is of the most important factors of heart diseases.

Diagnosis of heart diseases is a significant and boring task and also an important duty in medical science, which requires extreme attention. However there is some tools for data extraction and analysis. Also existence of huge set of medical data leads to correct diagnosis of disease. Using medical data including age, sex, blood pressure, and blood sugar, it is possible to increase the possibility of heart diseases prediction. These data must be collected in organized manner, which could be used for integrating the prevention system [4].

In different countries using artificial intelligence techniques and various algorithms, predicting this type of death -heart disease- is somewhat possible. In Iran many efforts in this subject made by cooperation of software and medical communities. So the current study focuses on field studies by ail of decreasing the cost and early prediction of events happened for heart patients in Iran.

One of important methods in this field is clustering. In clustering, the data splits to some clusters, in such a way that the data in every cluster have maximum similarity with each other and minimum similarity with data of other clusters. So using clustering data will show that every cluster that has the patient, could help us in predicting that if he/she is under heart attack risk or not. Using this method and reaching to more precise diagnosis for heart patients is our aim.

We know that traditional clustering methods like K-Means often judges on data using the distance between them. But in this study the highest objection is using this property, because available data about heart patients [5] includes binary and nominal data. So using various improved clustering methods, we can use other metrics [6] instead of focusing on distance between data, to focus on qualitative properties, to increase the precision and gain more correct diagnosis. By comparing and clustering the techniques and algorithms which used in heart diseases field and its diagnosis, which showed in Table (1), we will see that there is no algorithm which always has maximum performance, and various factors are effective in performance of algorithms.

Table 1. Comparing past works

Authors/Year	Used algorithm	Used Data Mining Method	Special Usage	Precision of selected Model
Mahmoudi/2013	CART Neural Network	Perceptron Neural Networks and Changing Selection based on Regression Tree and Clustering	coronary artery disease (heart attacks)	Neural Network: 74.19% CART 74.19%
Chaurasia et al. /2013	CART ID 3 DT (decision table)	Using three clustering algorithm	Early prediction of heart diseases	CART: 83.49%
Chaurasia et al. /2014	Bagging Naive Bayes J45DT	Using three clustering algorithm	prediction of heart diseases	Bgging:85.03 %
Ahmedet al. /2014	Decision Tree (DT), SVM, Neural Network	Using Clustering Methods	prediction of heart diseases	DT:92.1% SVM:91.0%
Dehghani et al. /2011	K-Means, Genetic Algorithm	Clustering based on combining genetic algorithm and K-Means for clustering data with common properties.	Diagnosis and predicting heart attacks	Performance based on calculated breaks: better determining clusters
Kurt et al. /2008	Logistic Regression, CART, Neural Network	Clustering techniques based on logistic regression, CART, neural network, and self-constructing map	Predicting coronary artery disease (heart attack)	Neural network: 78.7%

As it could be seen in Table (1), various studies on various algorithms with various methods and different data sets with special applications and even data with different types, carried out. As a result of evaluation and clustering carried out for diagnosis of coronary artery disease using data mining, this fact shows that never it is possible to introduce an algorithm by taking attention to its internal structure, as an optimized algorithm, and some other factors are effective in performance of these algorithms. In addition to selecting sub-set of properties as an effective factor, another effective factor was focused, which not only made changes in performance of algorithms, but also results in contradictory results.

Generally as the Figure (1) shows, there is several operational stages in preparing a model, which could be used as software application in predicting heart disease, as follows:

1) Selecting Data: First we must use existing database for creating models.

2) Removing Data: Remove incorrect and outlying data using statistical methods.

3) Integrating Data: Collect data from various resources into one general data-set.

4) Converting Data: All existing data must be normalized and converted to one format. This could be done using minimum and maximum normalizing methods.

5) Modeling: Creating model is main goal of a data mining project. In this stage some complex techniques and analysis methods used for extracting knowledge and information from data-set.

6) Clustering: The diagnosis of heart patient carries out using a combination of Fuzzy clustering algorithm and genetic algorithm, to gain more precise diagnosis in this disease.
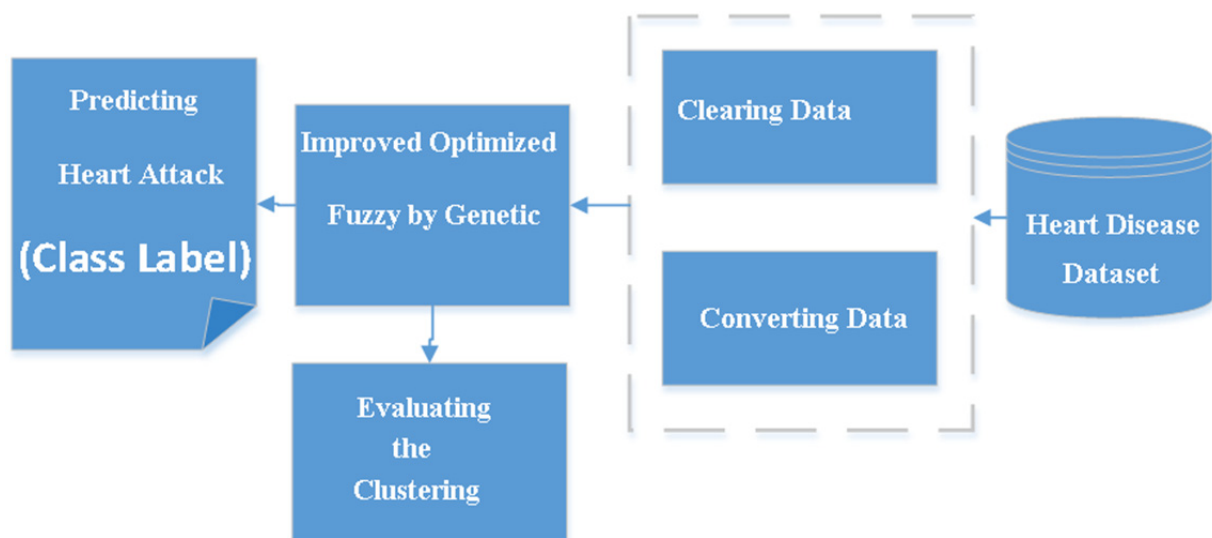

Figure 1. The Provided Methodology

This article contains five sections. The first section includes the problem definition and goal of this article and its history. In section two, the data-set of this article evaluated. Because the focus of this article is on clustering and prediction of heart disease, in section three a new algorithm for clustering and predicting heart disease introduced. Section four contains a summary and findings. In section five the conclusion, innovation and future suggestions about the subject of this article provided.

2. Data-Set

Data-sets of heart disease collected from 4 sources including:

I.   Cleveland Clinic

II. Hungarian Institute in cardiovascular diseases

III. Long-Beach medical center

IV. Medical Sciences University of Switzerland

All data of this database have the same format. This database has 76 attributes, but only 14 attributes used. All the attributes have numerical value and data-sets were used. Data-sets contain the following attributes:

1) Id: Identification number of patient.

2) Age

3) Sex: 1=man, 2=woman

4) Type of pain in chest (4 values)

5) Trestbps: Blood pressure.

6) Chol: Cholestrol of blood (mg/dl)

7) Famhist: Family history.

8) Restecg: Radiography records.

9) FBS: Fasting blood sugar (>120 mg/dl)

10) ECG: Electrocardiogram (value: 0, 1, 2)

11) Thalach: Maximum heart rate.

12) Pain in heart caused by sport.

13) Oldpeak: ST Depression caused by exercises.

14) Thaldur: ST section in inclined sport.

15) Flourosopy: The number of colored large vessels (3-0)

16) Thalassemia: 3=normal, 6=fixed defects, 7=reversible defects.

17) Num: Diagnosis of heart disease (Angiography status)

      a. Value 0: Possibility of narrowing the diameter of vessel is less than50% (healthy).

      b. Value 1: Possibility of narrowing the diameter of vessel is more than50% (patient).

From the 17 describes attributes, the attribute No. 1 is identification attribute and attribute No. 17 is the class title. Generally 14 attributes uses in diagnosis.

3. Methodology

The goal of this article is obtaining the most important factors and indexing heart patients from all patient's information, which has a vital role in heart disease for individual. So for doing this, we must help to precisely diagnosing the disease by increase the precision, robustness and performance of disease diagnosing in medical community, to bypass inefficient patient's information which is useless in disease diagnosis. Therefore in this article using multi-objective Fuzzy clustering techniques, a new method for diagnosing heart diseases provided, which by taking attention to the fact that qualitative data of database taken as the basis, we hope to help to medical community and also people. Using clustering algorithm based on nominal data with multi-objective fuzzy approach, which using genetic algorithm tries to optimize the internal goals of clusters, made patients grouping a more precise task and differentiates this work from previous ones.

Assume that $X = \{x_1, x_2,...,x_n\}$ is a set of n objects in discrete categorical zone. Every object, x_i , i=1, 2,…, n, describes by a set of p properties: $A_1, A_2,...,A_p$. Assume that $DOM(A_j)$, $1 \leq j \leq p$ is a domain of jth property and includes q_j different groupings, such that $DOM(A_j) = \{a_j^1, a_j^2,...,a_j^{q_j}\}$. So ith discrete object defined as $x_i = [x_{i1}, x_{i2},...,x_{ip}]$ which $1 \leq j \leq p, x_{ij} \in DOM(A_j)$. The center of cluster in FCM supersedes by modes of every cluster in Fuzzy K-Mode clustering. The defined modes are as follows: assume that C_i is a set of discrete objects which belongs to ith cluster. The mode of C_i is a vector, $m_i = [m_{i1}, m_{i2},...,m_{ip}]$, which $1 \leq j \leq p, m_{ij} \in DOM(A_j)$, in such a way that minimize the following criterion:

$$D(m_i, C_i) = \sum_{x \in C_i} D(m_i, x)$$

Where $D(m_i, x)$ is the amount of dissimilarity between m_i and x. Necessarily m_i is not one of the members of C_i set. Fuzzy K-mode clustering algorithm is on data x with K clusters for minimizing the following criterion:

$$j_m(U, Z : X) = \sum_{k=1}^{n} \sum_{i=1}^{K} u_{ik}^m D(z_i, x_k)$$

There is some conditions for probabilistic fuzzy clustering as follows:

$$0 \leq u_{ik} \leq 1, \quad 0 \leq i \leq K, \quad 0 \leq k \leq n$$

$$\sum_{i=1}^{K} u_{ik} = 1, \quad 1 \leq k \leq n$$

and

$$0 \leq \sum_{k=1}^{n} u_{ik} \leq n, \quad 1 \leq i \leq K$$

The fuzzy power is m, and $K \times n$ fuzzy clustering matrix is $U = [u_{ik}]$, and k^{th} discrete object's membership degree in i^{th} cluster is u_{ik}. $Z = \{z_1, z_2, ..., z_K\}$ Shows the centers of clusters (modes).

The Fuzzy K-mode algorithm is a part of periodic optimizing strategy, which contains the repetition of estimate of clustering matrix, and calculating the new cluster centers (modes). It starts with K random primary modes, and then in each repetition, the fuzzy membership on every data point in every cluster calculates by the following equation:

$$u_{ik} = \frac{1}{\sum_{j=1}^{K} \left(\dfrac{D(z_i, x_k)}{D(z_j, x_k)} \right)^{\frac{1}{m-1}}}, \quad 1 \leq i \leq K, \quad 1 \leq k \leq n$$

If $D(z_j, x_k)$ become zero for some j, so $u_{ik} = \{0, ..., 0\}$ for $i = 1, ..., K$, $i \neq j$ and for $i = j$ and $u_{ik} = 1$. Based on membership values, the cluster centers (modes) recalculates as follows. If the membership value was fixed, then the condition of modes which minimize the goal function are as follows:

$$z_i = [z_{i1}, z_{i2}, ..., z_{ip}]$$

in such a way that $z_{ij} = a_j^r \in DOM(A_j)$ and

$$\sum_{k, x_{kj}=a_j^r} u_{ik}^m \geq \sum_{k, x_{kj}=a_j^t} u_{ik}^m, \quad 1 \leq t \leq q_j, \ r \neq t$$

The finish condition of algorithm is when there was no significant improvement in J_m value. Finally, every object assigns to a cluster in which has maximum membership value in it. The main disadvantages of Fuzzy K-mode clustering algorithm are as follows:

1. This algorithm highly depends on choosing primary modes.

2. This algorithm often trapped in local optimal. So there is a need for an evolutionary optimization algorithm.

3.1 Displaying Chromosome

Every chromosome is a sequence of property values for using K mode of every cluster. If every discrete object

has p properties, $\{A_1, A_2, ..., A_p\}$, then the length of chromosome will be $K \times p$, in which the first location of P

(gene) shows the P-dimension of first cluster's mode, the second location of P shows the second cluster's mode,

etc. Assume that P=3 and K=3. So the chromosome is as follows:

Which shows three cluster modes including: $(c_{11}, c_{12}, c_{13}), (c_{21}, c_{22}, c_{23}), (c_{31}, c_{32}, c_{33})$. c_{ij} is the jth property of

i[th] cluster mode, and $c_{ij} \in DOM(A_j), 1 \le i \le K, 1 \le j \le p$.

3.2 Primary Population

K primary cluster modes coded in every chromosome and selected as K random objects from discrete data-set.
This process repeats for every one of chromosomes of population.

3.3 Calculating the Evaluation Function

In this study, there is two global compressing and resolution functions, which are two goal functions which
considered parallel. First the coded modes extracted, and are as $z_1, z_2, ..., z_K$.

This means that a level of A_j property from center of cluster, z_i, is a set of values which maximizes the sum of
u_{ij} (membership degree of ith cluster). Based on it, the value of cluster membership recalculates. The diversity,
σ_i and fuzzy cardinality, n_i in i[th] cluster which $i = 1, ..., K$ calculates with the following equation:

$$\pi = \sum_{i=1}^{K} \frac{\sigma_i}{n_i} = \sum_{i=1}^{K} \frac{\sum_{k=1}^{n} u_{ik}^m D(z_i, x_k)}{\sum_{k=1}^{n} u_{ik}}$$

For calculating the fuzzy separable suitability function, Sep, assume that z_i mode on i[th] cluster is center of a

fuzzy set, $\{z_i \mid 1 \le j \le K, j \ne i\}$. So the membership degree of every z_j toward z_i which $j \ne i$, calculates as

follows:

$$\mu_{ij} = \frac{1}{\sum_{l=1, l\ne j}^{K} \left(\frac{D(z_j, x_i)}{D(z_j, x_l)} \right)^{\frac{1}{m-1}}}, i \ne j$$

So the fuzzy resolution defined as follows

$$Sep = \sum_{i=1}^{K} \sum_{j=1, j\ne i}^{K} \mu_{ij}^m D(z_i, x_j)$$

It must be remembered that the compression of clusters became possible with π minimization. In return, for

obtaining a good separated cluster, the fuzzy resolution, Sep must maximized. So the above two criterion

considered as two goal of optimization functions in evolutionary algorithm. So we try to minimize the π and

$\frac{1}{Sep}$ with evolutionary algorithm.

3.4 Selection, Cross-Over and Mutation

The goal of multi-objective clustering is simultaneous optimization of more than one goal and increasing the

performance, using selection of these goals. The precise selection of goals could provide acceptable results, and foolish selecting goals provides bad results.

In this study we used common genetic operators. The operators used include: binary tournament selection operator, single point cross-over operator and single-point mutation operator.

4. Discussion

In this section, by taking attention to the provided algorithm for classifying heart patients, and taking attention to the heart patients' data, the Table (1) shows some samples of heart data.

Table 1. Several samples of heart data

	Age	Sex	Paint Type	Blood Pressure	Cholesterol	Sugar	Radiography	Heart Beat	Angina	Exercise	Inclined Sport	Number of Vessels	Blood Type
Sample 1	70	1	4	130	322	0	2	109	0	2.4	2	3	3
Sample 2	67	0	3	115	564	0	2	160	0	1.6	2	0	7
Sample 3	57	1	2	124	261	0	0	141	0	0.3	1	0	7
Sample 4	64	1	4	128	263	0	0	105	1	0.2	2	1	7
Sample 5	74	0	2	120	269	0	2	121	1	0.2	1	1	3

In this section, to providing the fuzzy clustering algorithm for predicting heart patients by running multi-objective fuzzy clustering based on genetic algorithm, an optimized Pareto front forms, meaning that finally we will gain a set of optimized final answers of solutions.
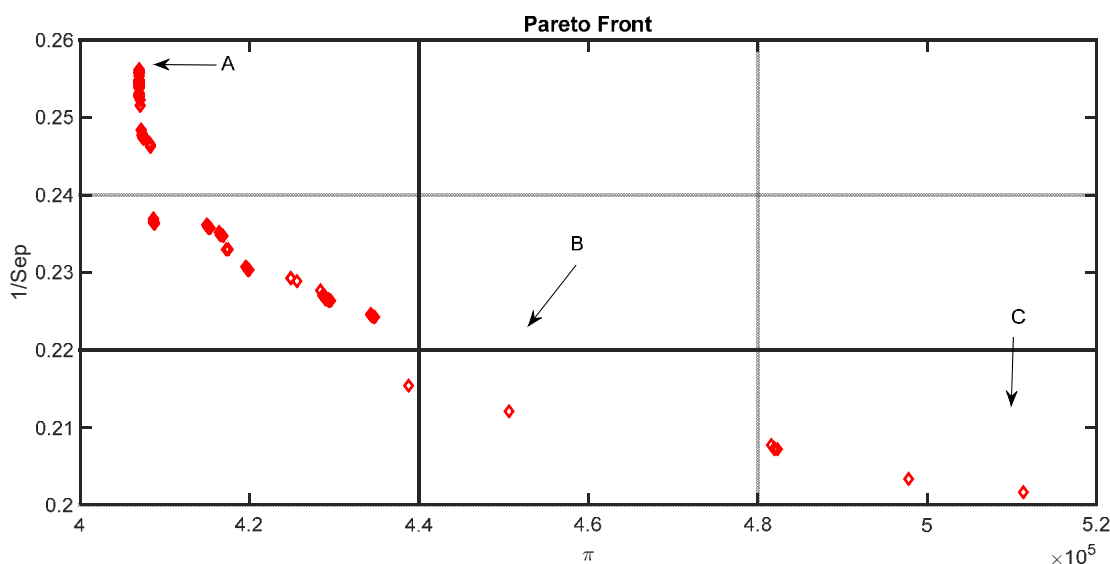

Figure 2. Optimized Pareto front

Based on Figure (2), for analyzing the optimized Pareto front of the above picture, three points A, B and C selected. The points in right of A have the best answer based on resolution, and as much as the points going toward top, the quality of answers based on resolution becomes better. The points which placed in B zone have the best answer based on compression. But the points in zone C are ones that based on both of compression and resolution have compelling quality.   In other words, these points have average compression and resolution. The elbow points in the above picture shows the best status of clustering. For validation of suggested algorithm, we used various criterions in evaluating the suggested clustering method.

4.1 Silhouette Validation Method

It defined based on the average distance between every sample of a cluster with all available samples of that cluster, and average distance of total current available samples in other clusters with a specific cluster. Based on this point of view, the amount of diversity and data correlation of data determined, which the maximum values of this index used for determining the number of optimal clusters.

$$S(i) = \frac{(b(i) - a(i))}{\max\{b(i), a(i)\}}$$

Where a(i) shows the non-similarity of a sample with other samples in a cluster and b(i) shows the non-similarity of a sample with all other samples of other clusters.

The value of Silhouette validation index is between -1 and +1. The higher values of this index (near to +1) shows that clustering made correctly. If index is near zero, means that we can assign a sample to a nearer cluster, and the sample is located in the same distance from both clusters. If the index becomes -1, it means that the clustering did not made correctly.

S(C) is an OSI (Overall Silhouette index) index for a clustering $C = \{C_1, C_2, ..., C_n\}$ and is the average value of Silhouette for all clusters:

$$S(C) = \frac{1}{n}\sum_{i=1}^{n} S(i)$$

The Silhouette index used for evaluating the level of similarity between suggested algorithm and heart patients classifying. Figure (3) shows that the results of Silhouette index explains the maximum similarity of formed clusters by suggested algorithm for classifying heart patients.
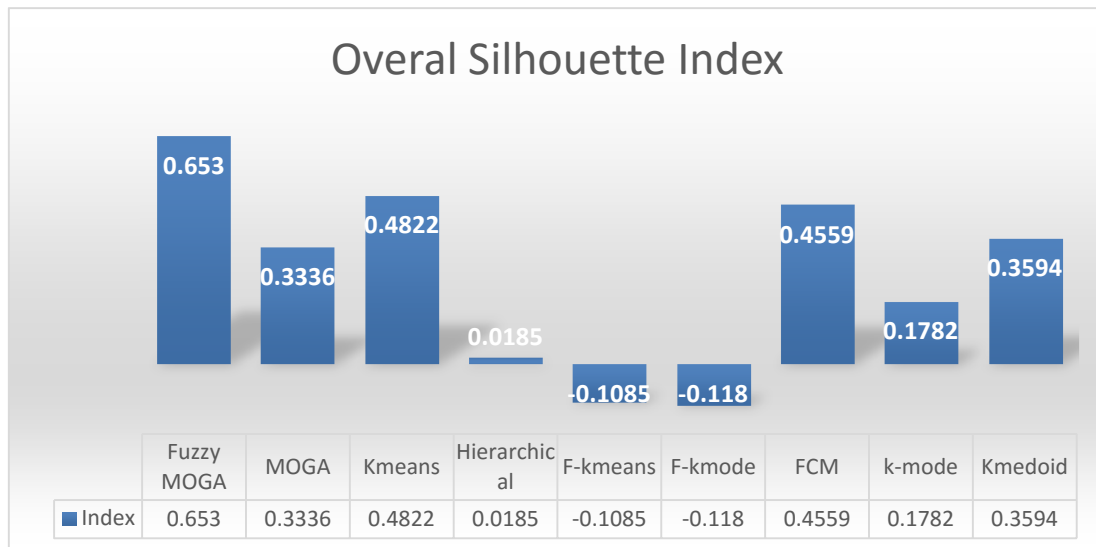

Overall Silhouette Index

	Fuzzy MOGA	MOGA	Kmeans	Hierarchical	F-kmeans	F-kmode	FCM	k-mode	Kmedoid
Index	0.653	0.3336	0.4822	0.0185	-0.1085	-0.118	0.4559	0.1782	0.3594

Figure 3. Comparison between Silhouette Index with other algorithms

4.2 Method of Validating the Dunn Index

This index defined by the following equation:

$$D = \min_{i=1,...,x_c}\left\{\min_{j=i+1,...,x_c}\left(\frac{d(c_i, c_j)}{\max_{k=1,...,x_c}\left(diam(c_k)\right)}\right)\right\}$$

In the above equation, d(x, y) and $diam(c_k)$ defined as follows:

$$d(c_i, c_j) = \min_{x \in c_i, y \in c_j} \{d(\mathrm{x, y})\}$$

$$diam(c_i) = \max_{x, y \in c_i} \{d(\mathrm{x, y})\}$$

If a data-set is separable, it is expected to have high distance between clusters and low diameter for clusters. As a result the higher value of this index is more favorable.

Table 2. Comparing Dunn index with various number of clusters

	Kmeans	Kmode	F-Kmode	Linkage	Fcm	Kmedoid	F-Kmean	MOGA	Fuzzy MOGA
2 clusters	0.000369	0.000105	0.000243	0.000865	0.000269	0.000116	9.39E-05	0.067	0.05857896
3 clusters	0.000712	9.06E-05	9.06E-05	9.84E-05	0.000903	0.000193	0.000272	0.0464	0.126
4 clusters	0.000188	6.39E-05	0.000274	0.000283	0.001	0.000318	0.000657	0.1376	0.073
5 clusters	0.000254	0.000106	0.000174	0.000649	0.000201	0.000129	0.000103	0.037	0.0529

Results of Dunn validation method showed in Table (2). Based on the results it specified that in suggested algorithm, the Dunn index has better and more suitable results in comparison with other provided algorithms. So the quality of made clusters were higher and better than others. The range of classifying data by suggested fuzzy clustering algorithm was 2-5 classes. In this article, based on Tables (3) and (4), by comparing results of Dunn index for every made clusters by suggested algorithm, the most optimized case made by 3 clusters of heart patients' data suggested.

Table 3. Obtained P-Values from test

	Kmeans	Kmode	F-Kmode	Linkage	Fcm	Kmedoid	F-Kmean	Proposed
2 clusters	1.7159e-55	7.4871e-52	6.9935e-61	1.5480e-69	1.1582e-53	7.2913e-66	3.92E-54	5.1680e-59
3 clusters	1.2802e-54	6.4155e-41	2.1380e-39	1.4290e-49	1.9016e-54	1.1580e-47	4.76E-50	1.7949e-57
4 clusters	1.4071e-53	4.2195e-36	7.7059e-40	2.1860e-79	4.9969e-48	1.4753e-55	1.67E-48	5.6419e-68
5 clusters	2.0972e-51	1.0741e-42	1.8236e-42	2.3334e-60	2.9709e-54	2.6559e-50	7.00E-52	5.1143e-51

Finally for evaluating the level of validity in results of every clustering algorithms we used T-test for every one of them in 5%significance level. P-value shows the probability level of chance in obtained information for Dunn index in every cluster. The P-values obtained from every algorithm shows the meaningfulness of results from the statistical point of view. Based on the zero hypothesis, we assume that there is no meaningful difference between values of Dunn index in any cluster. While the other hypothesis says that there is meaningful difference between them. Based on Table (3) all of obtained P-values are less than 0.05. So the obtained results are precise. The Figure (4) shows the comparison between Pareto Front with number of different clusters.
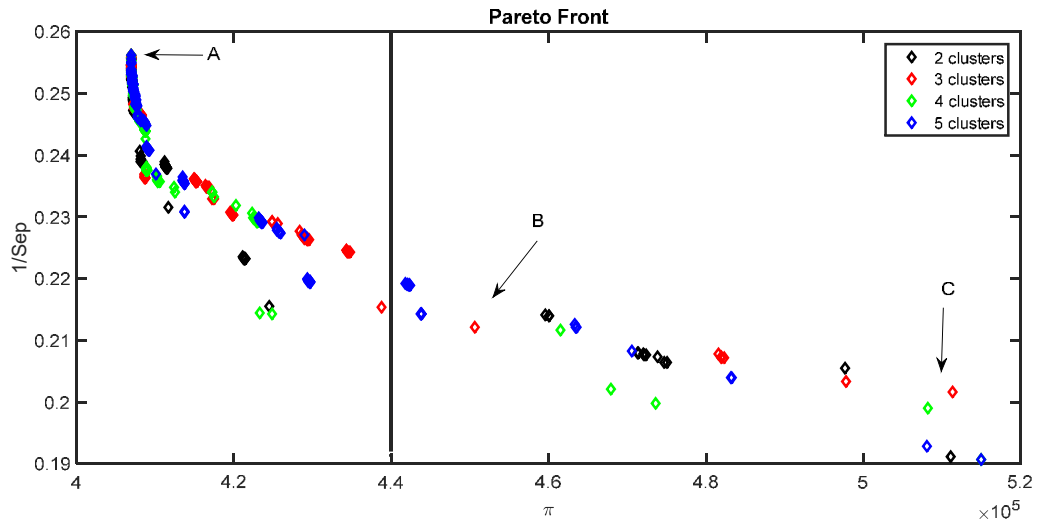
Figure 4. Comparison between Pareto Front with different clusters

Figure (5) shows the relationship between different classes of probability of heart attack with age and sex of patients. For example patients of cluster 2 are mainly men that will be suffering heart attack. Figure (6) shows the relationship between different classes of heart attack probability with age and sex of patient. By studying on second cluster we can evaluate and diagnose the signs of heart diseases.
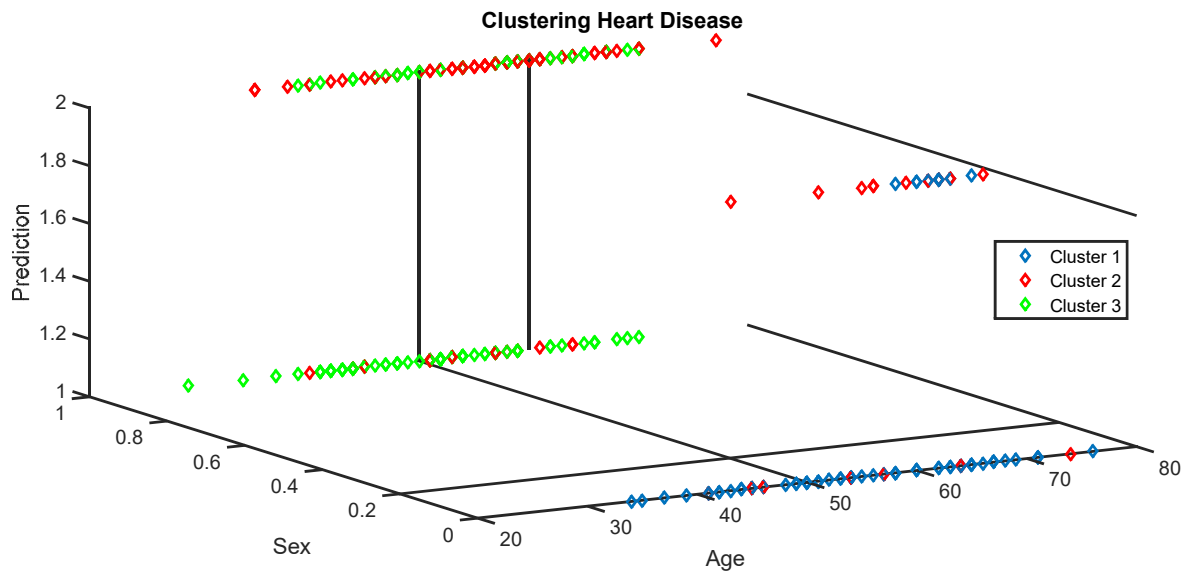
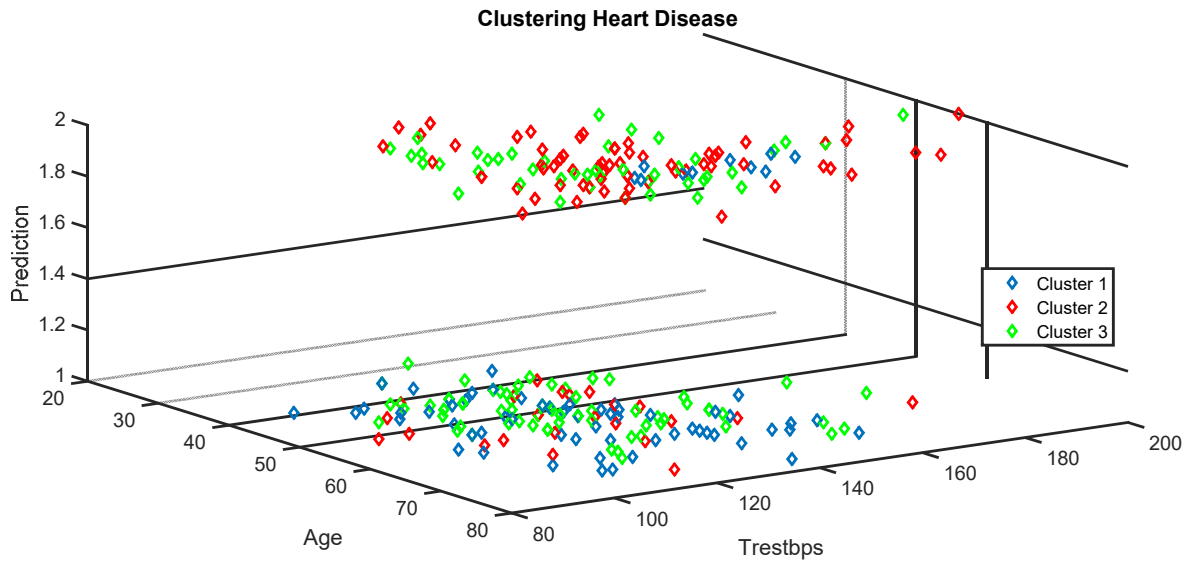Figure 5. Relationship between different classes of heart attack probability with age and sex of patient

Figure 6. Relationship between different classes of heart attack probability with age and blood pressure of patient

In Figure (7) the relationship between different classes of heart attack probability with age and sex of patient is obvious. In this figure we can see that by studying the second cluster –colored with red- we can diagnose the signs of heart attack and use it for later diagnoses. Figure (8) shows the relationship between different classes of heart attack probability with blood pressure and cholesterol of patient. Also in this figure it is obvious that the second cluster specifies the heart attack probability.
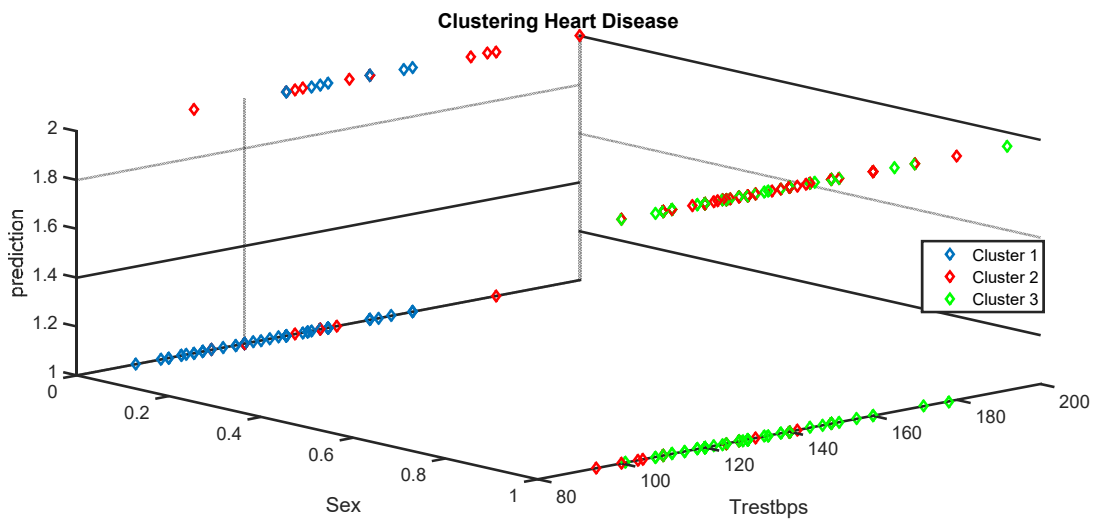

Figure 7. Relationship between different classes of heart attack probability with blood pressure and sex of patient
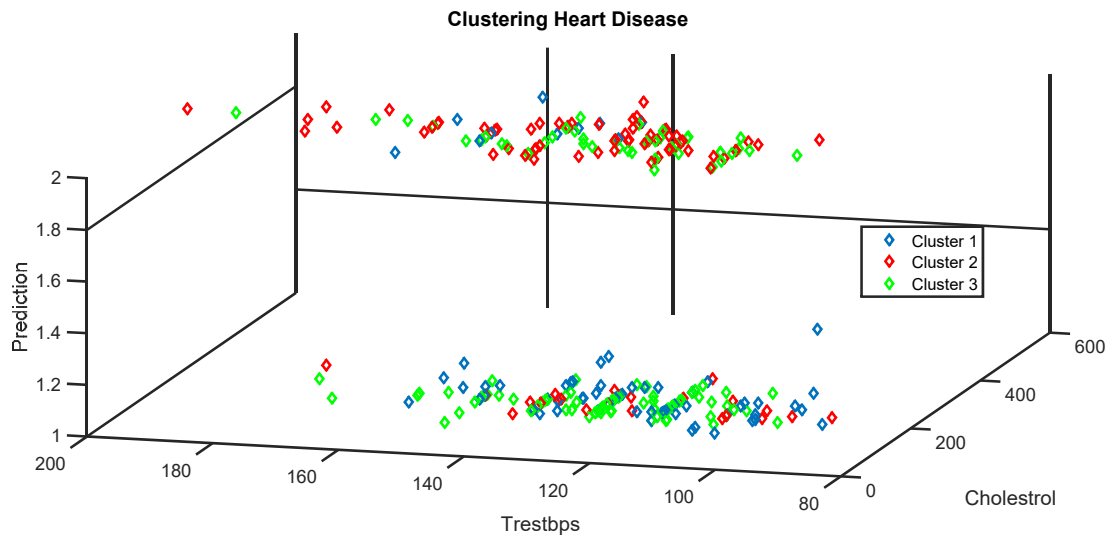
Figure 8. Relationship between different classes of heart attack probability with blood pressure and cholesterol of patient

5. Conclusion

Heart is a pump or pulsating pump which composed of four compound holes with two atriums and two ventricles, which delivers blood to all body organs. So the heart is a vital organ of body. Unfortunately todays an important quantity of death is caused by heart diseases. Todays the cardiovascular diseases are the most important challenges of healthcare in the worldwide. Prevention and management of cardiovascular diseases requires a pervasive and comprehensive system for recording data. Information of patient records are one of the most important data, which must be classified for easy and fast treatment process. Goal of current study was providing a classifying system for cardiovascular diseases to improve the policies of healthcare in opposition against cardiovascular diseases. The main goal in classifying is putting people in groups with predetermined number of patients. Then the future patients will be helped with evaluating their signs.

In all developed countries the diseases classifying system is a fundamental base for addressing healthcare requirements of country with classifying cardiovascular diseases. The national system of classifying cardiovascular diseases will play an effective role in improving the management and prevention of cardiovascular diseases in Iran. Using this approach we can greatly help in early prediction of this disease.

In this article we introduced some of the most useful algorithms and techniques of artificial intelligence which recently used, and briefly described their properties. In recent years several studies carried out in artificial intelligence subject on heart patients' data, and many of algorithms were successful which mentioned earlier. But the important point is that the level of success in these algorithms depends on various factors and it is not possible to choose a method as the best one. Factors like data type of database, selecting sub-set of properties and risk factors, number of properties, the larger size of database, low number of missed data and access to suitable and correct data increases the success chance in exploration and increase the quality of algorithms' results. In this subject we could mention to cases which improves the algorithm including: using self-adaptive neural network which obtained from fuzzy clustering, or using a combination of clustering and SVM classifying, and using fuzzy rules discovery on heart patients data.

References

Ahmet, M., Niyazi, K., & Aydin, A. (2014). Evaluation of bagging ensemble method with time-domain feature extraction for diagnosing of arrhythmia beats. *Neural Computing and Applications, 24*(2), 317-326.

Dehghani, T., Afshari, S. M., & Khalilzadeh, M. (2011). *A genetic K-means clustering algorithm for heart disease data*. 5 th Conference of Data Mining of Iran, Amirkabir University.

Imran, K., Mevlut, T. A., & Turhan, K. (2008). Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. *Expert Systems with Applications, 34*, 366–374.

Jesmin, N., Tasadduq, I., Kevin, S. T., & Yi, P. C. (2013). Association rule mining to detect factors which contribute to heart disease in males and females. *Expert Systems with Applications, 40*(4), 1086-1093. Retrieved from http://archive.ics.uci.edu/ml/machine-learning-databases/statlog/heart/heart.dat

Mahmoudi, B., Pohlmeyer, E. A., Prins, N. W., Geng, S., & Sanchez, J. C. (2013). Towards autonomous neuroprosthetic control using Hebbian reinforcement learning. *J. Neural Eng. Dec., 10*(6), 066005.

Mourão, A., Martins, F., & Magalhães, J. (2015). Multimodal medical information retrieval with unsupervised rank fusion. *Computerized Medical Imaging and Graphics, 39*, 35–45.

Robert, O., Bonow, D. L. M. M., Douglas, P., Zipes, M. D., & Peter, L. (2012). Braunwald's Heart Disease A Textbook of Cardiovascular Medicine.

Roohallah Alizadehsania, J. H., Mohammad, J. H., Hoda, M., Reihane, B., Asma, G., Behdad, B., & Zahra, A. S. (2013). A data mining approach for diagnosis of coronary artery disease. *Computer Methods and Programs in Biomedicine, 111*(1), 52–61.

Vikas, Ch., & Saurabh, P. (2013). Data Mining Approach to Detect Heart Diseases. *International Journal of Advanced Computer Science and Information Technology (IJACSIT), 2*(4), 56-66.

Vikas, Ch., & Saurabh, P. (2013). Early Prediction of Heart Diseases Using Data Mining Techniques. *Carib. j. SciTech.*, 208-217.

Yang, C. L. et al. (2015). Non-dominated sorting genetic algorithm using fuzzy membership chromosome for categorical data clustering. *Applied Soft Computing, 30*, 113–122.

Copyrights