IJE TRANSACTIONS B: Applications Vol. 30, No. 8, (August 2017) 1152-1159

International Journal of Engineering

Journal Homepage: www.ije.ir

An Efficient Predictive Model for Probability of Genetic Diseases Transmission

H. Hamidi*[a], F. Qaribpour[b]

Department of Industrial Engineering, Information Technology Group, K. N.Toosi University of Technology, Tehran, Iran

PAPER INFO

Paper history:
Received 05 May 2017
Received in revised form 18 June 2017
Accepted 07 July 2017

Keywords:
Genetic Disease
Transmitting Genetic Diseases
K-means Clustering Algorithm
Data Mining
Decision Trees

ABSTRACT

In this article, a new combined approach of a decision tree and clustering is presented to predict the transmission of genetic diseases. In this article, the performance of these algorithms is compared for more accurate prediction of disease transmission under the same condition and based on a series of measures like the positive predictive value, negative predictive value, accuracy, sensitivity and specificity. The results show that support vector machine algorithm outperformed the other two simple algorithms and the neural network and genetic algorithms offered better prediction at the end, while the proposed combined approach is developed using different parameters and outperformed the simple methods.

***doi**: 10.5829/ije.2017.30.08b.06*

1. INTRODUCTION

The medical field is expanding and evolving every day and constantly producing large amounts of data [1, 2]. Medical data are produced in more different ways than in the past [3, 4]. The need for an efficient and accurate solution for new managements makes more sense than ever before [5-7]. In this regard, there is need for methods and automated algorithms to operate this volume of data with different forms so that we can analyze the data, then perform operations on them and to achieve the desired results [8-10]. So, smart and flexible methods and data mining algorithms, which are appropriate to the data and their changes, are needed [11, 12] and genetic data are not exempt from these data. With regard to genetic data, there is also need for an approach that in addition to reviewing the data, analyze them as well [13-15]. One of the main problems for choosing this approach is that providing genetic data for each patient may requires the need to save characteristics that is not needed for other patients; for example, There may be the need to save a patient's blood test results but for other patients it is not a priory to do this test and save its or we may encounter cases that were initially not foreseen while examining the patient's condition. For this reason, it is better not to design a general plan for the database from the beginning so that we have the possibility of adding any characteristic that is needed during the operation. Another important issue is that considering the need to genetic data of family members and the previous generations, it must be possible to add new attributes (previous and next generations) while performing researches in this database [16, 17]. Genetic diseases are one of the main causes of failure in the world and early diagnosis and prevention is the best treatment. Data mining can be effectively used in the fast and cost-effective prediction and diagnosis of diseases [16, 18, 19]. In this regard, considering the volume and format of data for the study of transmission of genetic disorders and importance of studying the relationship between individuals in this type of disease, it is important that each patient may require specific data exist, thus it is not possible to define the predetermined schema. With regard to genetic diseases, data which are required to be saved are varied. Considering the nature of genetic disease, there is also the need to save health status of patients' ancestors to understand the transmission of these diseases and new person may be added to the family tree in each investigation. Also, it is very

*Corresponding Author's Email: *h_hamidi@kntu.ac.ir* (H. Hamidi) important to explore the transmission route and the relationships between individuals in this database [20-25]. So far, few models have been used to analyze medical data but each of these data model has a number of disadvantages which make them to be non-ideal data model. One of these data model is the relational data model. The next model is the object-relational data models, which solve the predefined schema problem and well define different data formats using Entity attribute value (EAV) design, but still remains the problem of relationships between individuals [17, 26]. The main objective of this article is to use data mining techniques for data clustering by K-means algorithm and combine it with the decision tree in order to predict the risk of transmitting genetic diseases. The following article is organized as follows: Review of the literature is presented in the second section. In the third section, the article methodology has been stated. Findings, discussion and analysis are later discussed in the fourth section. Final conclusions are dealt with in the fifth section.

2. REVIEW OF THE LITERATURE

Today, large volumes of medical data are generated in various forms. Efficient analysis and use of this data in a reasonable time requires appropriate data mining approaches. The purpose of this article is to study the genetic data in order to predict the likelihood of transmission of genetic disorders. Xu et al. [6] conducted a survey research on data mining approaches used to predict heart diseases. They stated that data mining is the process of finding useful and relevant information from the database. There are several types of data mining techniques. Association rules, classification, neural networks, clustering are among the most important data mining methods. Data mining process plays an important role in industry and health service. The data mining processes are widely used in the healthcare field to predict diseases. In their paper, they analyze different types of processes to predict heart diseases using data mining [27-29]. Daraei et al. [7] investigated and analyzed the use of association classifiers in the predictive analysis in data mining in the field of health and medical services. Association rule mining is one of the most important data mining techniques for anatomical tasks and a lot of researches has been conducted in this area and was used for the analysis of tool basket. Classification using association rules is another primary predictive analysis method, which aims to discover small set of rules in the mass of data of databases which are considered an accurate classifier. They expressed that they have offered a combined approach in this article that combines association rule mining with classification rules mining and call it association classification (AC). This is a new classification approach. This integration was carried out with a focus on mining a particular subset of association rules that are called association rules classify (CAR) and classification is carried out by these CARs. The use of association rules mining for classification systems is a promising approach. Considering their legibility, association classifiers are very useful and convenient for experts in the decision-making process. The medical world is a good example of such application. Consider a situation, in which a physician wants to examine a patient. There is vast amount of information about the patient's condition (Including personal data, test results, etc.). A classification system can help the physician in this work. The system can assess whether the patient is at risk of certain diseases in the future or is incompatible with some treatments. Given the output of the classification model, the physician can make a better decision regarding the patient treatment. Combining advanced classification rules mining with classifiers provides a new type of association classifiers. They will discuss that this advanced association classifiers, which have been proposed in recent years, are more accurate than traditional classifiers [30-32]. Stork et al. [14] investigated the use of decision tree to diagnose the heart disease of patients. They stated that the heart disease is the leading cause of death in the last 10 years. The researchers used several data mining techniques in order to help healthcare professionals diagnose heart diseases. The decision tree is one of data mining methods which have been successfully used in recent years. However, most researches have used the J 4.8 decision tree based on interest rates and binary discretization. Gini index and information gain are two other successful decision trees that have been less used in the diagnosis of heart disease. Also, other discretization techniques, voting method, and reduced error pruning were known as better decision trees. This research investigates the application of some techniques in different types of decision trees to obtain better performance in the diagnosis of disease. Bench-marking databases which are widely used are used in the research. To evaluate the performance of alternative decision trees, sensitivity, specificity and accuracy were calculated. This research proposes a model, in which J 4.8 decision tree and bag algorithm has a better performance in the diagnosis of heart disease [7, 33-37]. In reference [6], the association rules and decision trees were evaluated to predict the multifunctional properties. It is stated that the association rules, decision trees and data mining techniques are well-known in finding predictive rules. In this study, they provided a detailed comparison on the association rules and decision tree so that the multifunctional properties are predicted and identify the important differences between the above two techniques for such purpose. They conducted an extensive test on the actual medical databases so that the data mining process is conducted on rules that predict a disease in several coronary arteries. The prerequisite for the association rules contains medical measurements and risk factors, while its consequences are degree of severity of the disease is in one or more arteries. Predictive rules by the association rules mining are more frequent and enjoy higher reliability than the predictive rules induced by the decision tree [38-41].

3. THE PROPOSED METHOD

In this article, the library method has been used in order to achieve the basic concepts and theoretical principles and definitions. Thus, resources, books, domestic and foreign articles, internet search, as well as the opinions of experts were used during its development in order to improve the quality of work and test assumptions.

Figure 1 shows the process of general implementation of the proposed approach. Firstly, considering the need for genetic data, genetic data, which are effective in transferring genetic disease, are collected. In the second stage, the criteria, which are priority in terms of importance and also in terms of the higher risk of diseases transmission, are separated from the collected data and in the third stage, priority data clustering is performed using K-means algorithm for the easier and more accurate data mining. The fourth stage relates to the decision-making about the fact that whether the individual with the existing characteristics transmitted genetic disease or not. Finally, the obtained results are evaluated in terms of accuracy and sensitivity and etc. of the proposed method in the last stage.

In this article, a collection of genetic data is used. There are many symptoms of genetic disease, finding patterns of the genetic disease data which are helpful in diagnosing the cause of the disease and its transmission in the future. Database comprised in this article consists of 303 samples, including 297 full samples and six samples with lost values. This database has 76 raw attributes while all trials were performed only on 5 of their attributes. So, this database contains 3 symptoms of the genetic disease transmission and the meaning of each of the symptoms will be described later (Table 1).

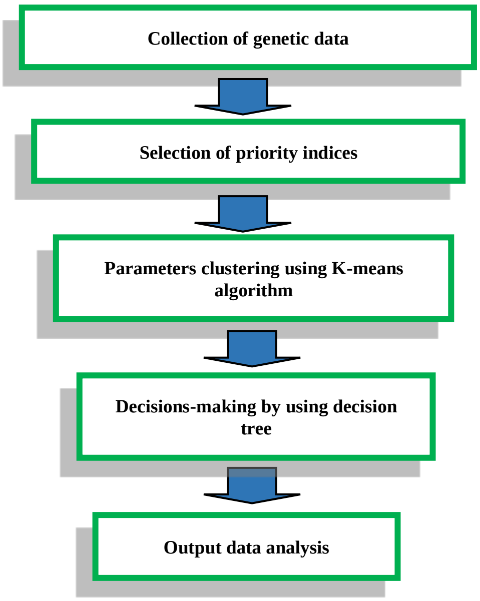
Collection of genetic data
Selection of priority indices
Parameters clustering using K-means algorithm
Decisions-making by using decision tree
Output data analysis

Figure 1. The process of implementing the proposed approach

Genetic diseases are divided into 3 categories: monogenic diseases, multi-genic or multi-factor diseases, mitochondrial diseases. Monogenic diseases are diseases that are transmitted from patient's parents or carriers to find children who suffer from the disease at birth, but their age of onset is different in the body, such as thalassemia. These diseases are divided into two dominant and recessive categories, which are described below. Multi-genic diseases are diseases, in which genes suffer from problems caused by environmental changes and disease occur in the body, such as cancers, in development of which eating and lifestyle habits like smoking play an important role. Mitochondrial diseases are diseases that are transmitted only from mother to child and not the father. There are two modes of transmission with regard to the monogenic diseases: recessive disease, dominant disease. Parents are not ill in the recessive disease and have no symptoms. However, their child may be infected with the disease from past generations. In fact, parents are only carriers of disease and are not ill in this type of disease, which includes albinism, skin cancers, cystic fibrosis, mental retardation, sickle cell anemia, phenylketonuria, thalassemia, etc. Mother of father or both suffer from the genetic disease and their children are very likely will be ill in the dominant disease such as dwarfism, multiple skeletal abnormalities, cataracts, muscle weakness and dystrophy, syndactyly, short toes with hands and feet disorders, psoriasis, Huntington's, cancer, eye retinas and etc. Other types of genetic diseases include hemophilia, Duchene muscular dystrophy (DMD), different types of deafness, cleft lip and cleft palate, mental illness, rickets resistant to vitamin D, favism, insipidus diabetes, color blindness etc. (gender-dependent diseases). A K-means algorithm is used in the first stage of data clustering process.

TABLE 1. Database components

Component	Label	Attribute
Patient's age	Age	Age
Patient's gender	Sex	Sex
Single-gene disease	SGD	Single-genic diseases
Multi-genic disease	MGD	Multi-genic diseases
Mitochondrial diseases	MtDNA	Mitochondrial diseases

It also increases the accuracy and speed of data processing. The proposed methodology which consists of different sections, the genetic database, includes a number of attributes, which are used to distinguish the risk of transmitting a genetic disease from the lack of transmission. As previously mentioned, the database contains 5 columns and 303 rows. 4 columns represent the attributes and 1 column represents the class label. The next step is to create a decision tree, through which the database components are weighted. A weighted mean is achieved for each person. Based on association rules, any person whose weighted mean is higher and less than 50% is at risk of disease transmission and safe, respectively. Figure 2 shows the step by step production of the decision tree.

In addition to having a simple and tangible structure and high accuracy, decision tree has an important attribute such as its feature selection. This means it specifies, among different entries, the entries that have more weight in the classification and are known as dominant feature. For example, in the following decision tree, among the various features used in the input, only four features, including single-gene diseases, multi-genic diseases, mitochondrial diseases and triggers of disease have been marked as dominant features affecting the classification. This property of the decision tree can be used in problems, the feature space of which is very large. Initially, the dominant features are selected using the decision tree and then the dominant features can be used as input of any custom expert system. It can improve the performance of the expert system. Figure 3 shows decision tree in case of manner of investigating data and decision-making.

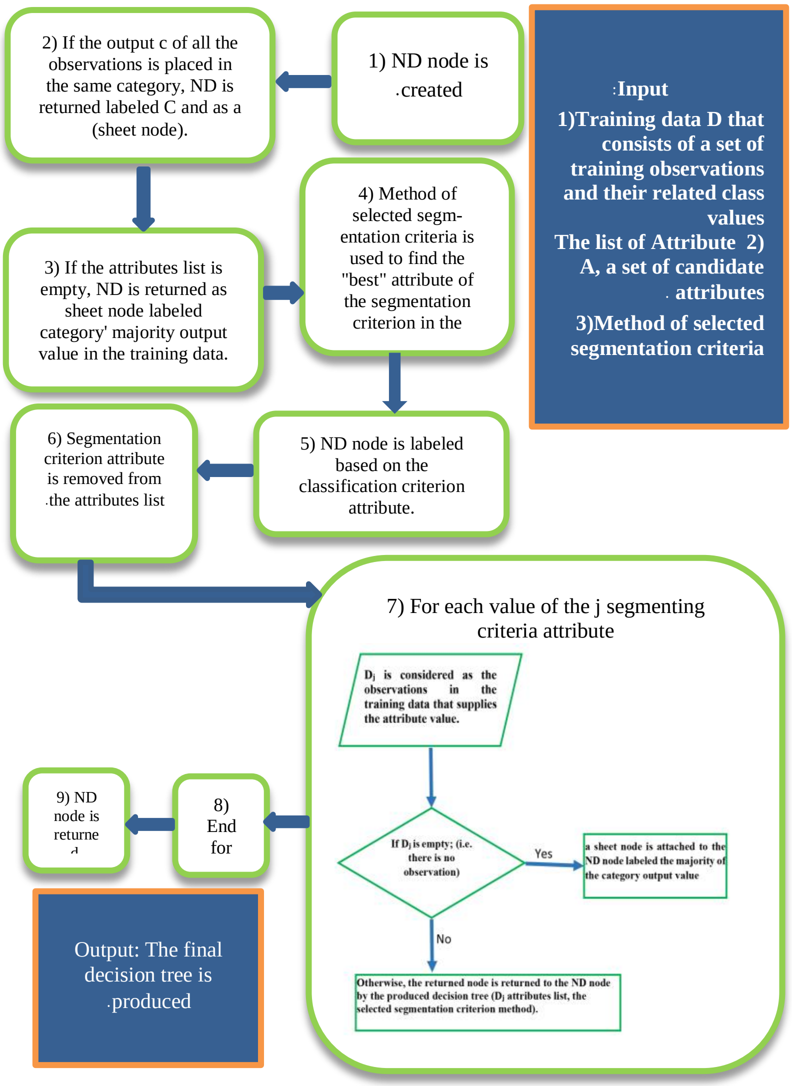
1) ND node is .created
2) If the output c of all the observations is placed in the same category, ND is returned labeled C and as a (sheet node).
:Input
1)Training data D that consists of a set of training observations and their related class values
The list of Attribute 2) A, a set of candidate . attributes
3)Method of selected segmentation criteria
3) If the attributes list is empty, ND is returned as sheet node labeled category' majority output value in the training data.
4) Method of selected segm-entation criteria is used to find the "best" attribute of the segmentation criterion in the
5) ND node is labeled based on the classification criterion attribute.
6) Segmentation criterion attribute is removed from .the attributes list
7) For each value of the j segmenting criteria attribute
Dj is considered as the observations in the training data that supplies the attribute value.
If Dj is empty; (i.e. there is no observation)
Yes
a sheet node is attached to the ND node labeled the majority of the category output value
No
Otherwise, the returned node is returned to the ND node by the produced decision tree (Dj attributes list, the selected segmentation criterion method).
8) End for
9) ND node is returne d
Output: The final decision tree is .produced

Figure 2. The production process of the decision tree

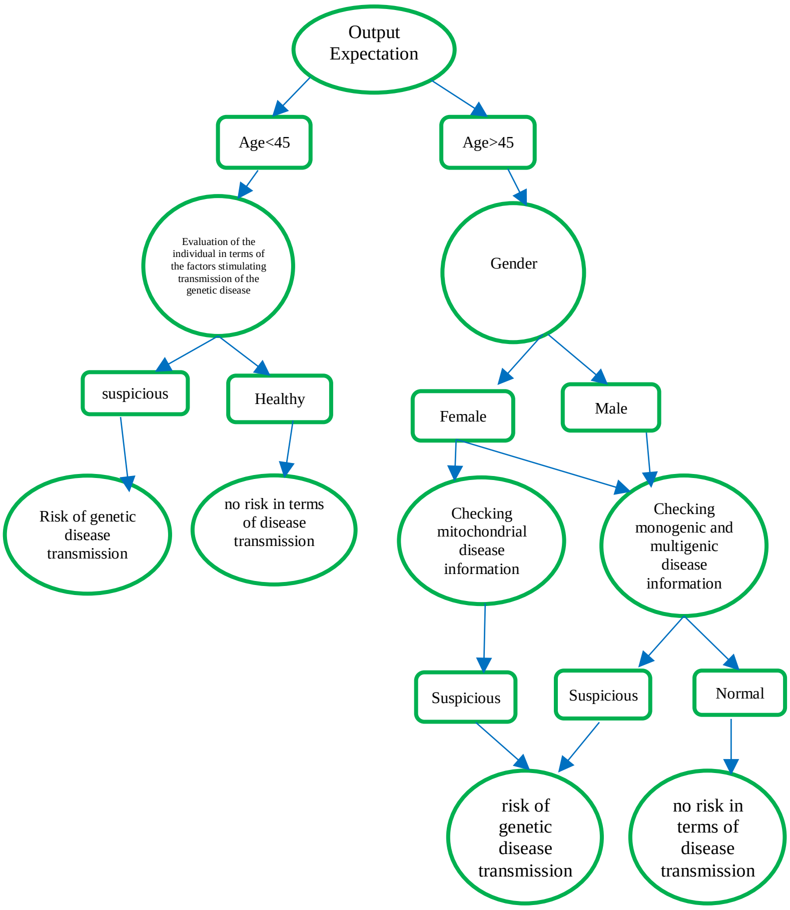
Output Expectation
Age<45
Age>45
Evaluation of the individual in terms of the factors stimulating transmission of the genetic disease
Gender
suspicious
Healthy
Female
Male
Risk of genetic disease transmission
no risk in terms of disease transmission
Checking mitochondrial disease information
Checking monogenic and multigenic disease information
Suspicious
Suspicious
Normal
risk of genetic disease transmission
no risk in terms of disease transmission

Figure 3. The decision tree of the proposed method

4. FINDINGS AND DISCUSSION

This method has been implemented by using the database and combining clustering algorithm and decision tree in MATLAB. The weight matrix that has been created for each person by weighting the decision tree based on the description of each of the components, is a matrix 5 × 303 and some part of it is given in Table 2 for 5 persons in database due to its high volume.

Now, the weighted mean matrix, which is 303 × 1 matrix is achieved for each individual based on the above table. Part of the above matrix is shown here due to lack of space (Table 3).

TABLE 2. Sample weight matrix

Mitochondrial disease	Multi-genic	Monogenic	Age	Gender	Individuals
3.772	0.256	1	0.313	0	**First person**
4.264	0.465	1	0.362	1	**second person**
4.201	0.125	0.6	1	0	**Third person**
3.602	0.075	0.7	1	1	**Fourth person**
4.083	1	0.6	0.235	0	**Fifth person**

TABLE 3. Sample weighted mean matrix

Criteria / algorithms	Accuracy	Specificity	Sensitivity	Negative predictive value	Positive predictive value
Genetic Algorithm	63	79	64	70	86
Neural network	71	80.5	76	78	87
Support Vector Machine	78	82.5	64.5	77	89.2
Proposed method	81	81.9	79	86	89

Individuals at risk can be identified using this weighted mean matrix using a decision tree. It's like this:

If the weighted mean is above 50%, the individual has high probability of transmission of the genetic disease. If the weighted mean is below 50%, the individual is not at risk of transmission of the genetic disease.

Different criteria are used to compare the results of the implementation of the proposed method and 4 other most commonly used algorithms:

1. *Accuracy:* Number of samples correctly diagnosed in the intended class compared with the whole samples.

2. *Sensitivity:* Number of samples which have correctly shown the absence of disease transmission compared to the total number of samples that have really not the genetic disease.

3. *Specificity:* Number of samples that have correctly shown the presence of disease transmission compared to the total number of samples that really suffered from the genetic disease

4. *The positive predictive value:* Number of samples that have correctly shown the absence of disease compared to the total number of samples which are not suffering from the predicted disease.

5. *The negative predictive value:* Number of samples that have correctly shown the presence of the disease compared to the total number of samples which are suffering from the predicted disease.

$$\text{Sensitivity} = TP / (TP + FN) \tag{1}$$

$$\text{Specificity} = TN / (TN + FP) \tag{2}$$

$$\text{Accuracy} = (TP + TN) / (TP + FP + TN + FN) \tag{3}$$

TP: The number of samples which are correctly diagnosed positive.

TN: The number of samples which are correctly diagnosed negative.

FP: The number of samples which are incorrectly diagnosed positive.

FN: The number of samples which are incorrectly diagnosed negative.

Based on the pervious assessments and their results, as is clear from Table 4 and Figure 4, if the proposed methodology is set aside and other common and simple method is compared, the following results are obtained: Vector machine algorithm is in the positive predictive value and superior feature and it has the lowest sensitivity and negative predictive value compared to the percentage of this criterion in genetic algorithm (GA). However, the proposed method has better performance in several criteria than other algorithms. Nevertheless, it is noteworthy that it is in the second place in terms of two criteria, including positive predictive value and specify, but in general, it is ranked higher than other methods especially based on the accuracy criterion, which partly led to better diagnosis of the genetic disease transmission.

TABLE 4. Comparison of different criteria on algorithms based on the percentage

Fifth person	Fourth person	Third Person	Second person	First Person	**Individuals**
0.612748	0.609258	0.464119	0.643598	0.633258	**Weighted mean**

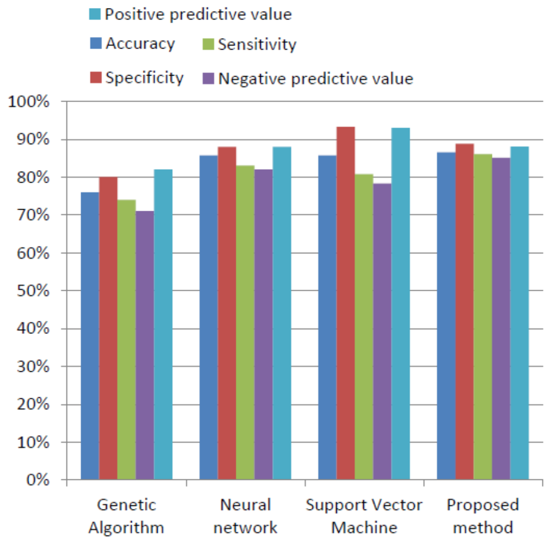

Figure 4. Diagram of the different criteria on algorithms based on the percentage

5. CONCLUSION

Faster and more efficient analysis of medical data requires automated and electronic analysis of the data. Data on genetic diseases are no exception of this set of data. Therefore, appropriate algorithm and methods of data analysis should be designed. So far, several methods have been used for analyzing medical data, but each of these methods has their own disadvantages which make them non-ideal approach. In this article, in addition to analyzing and evaluating the best data mining algorithms used in the medical field, a new combined approach has been provided in order to predict the risk of transmitting genetic diseases. Considering the nature of genetic data and the fact that these relationships between individuals and analysis is considered the major issue with regard to the transmission of genetic diseases, combination of K-means clustering methods and a decision tree is used in this article to analyze the data. At first, patients' data are clustered and then decision tree-based cluster analysis was performed to determine whether the person is likely to have a genetic disease transmission, or is safe. For this purpose, a database, including 303 samples, involving 297 complete samples and six samples with lost values has been formed. This database has 76 raw attributes while all trials were performed only on 5 of their attributes. The results of the patients clustering were obtained using the risk of transmitting genetic diseases and according to the criteria of similarity in the transmission ways as well as using a decision tree to predict whether the individual with the related characteristics is likelihood to transmit the disease or not. The study of obtained outputs while using this combined approach suggests that the proposed method has higher accuracy and sensitivity compared to other data mining algorithms.

Suggestions for future research:

- Other multi-criteria decision-making methods can be used in the future research to rank the worst patient.
- Future research could investigate direct and indirect effects of each of the factors affecting the choice of indicators.
- The use of the neural networks and fuzzy logic to predict the transmission of diseases.
- Combining meta-heuristic algorithms with the proposed method.

6. REFERENCES

1. Zhu, L., Wu, B. and Cao, C., "Introduction to medical data mining", ***Sheng wu yi xue gong cheng xue za zhi= Journal of biomedical engineering= Shengwu yixue gongchengxue zazhi***, Vol. 20, No. 3, (2003), 559-562.
2. Berka, P., "Data mining and medical knowledge management: Cases and applications: Cases and applications, IGI Global, (2009).
3. Ting, S., Shum, C., Kwok, S.K., Tsang, A.H. and Lee, W., "Data mining in biomedicine: Current applications and further directions for research", ***Journal of Software Engineering and Applications***, (2009).
4. Robinson, I., Webber, J. and Eifrem, E., "Graph databases: New opportunities for connected data, " O'Reilly Media, Inc.", (2015).
5. ŢĂRANU, I., "Data mining in healthcare: Decision making and precision", ***Database Systems Journal BOARD***, (2016), 33-40.
6. Xu, D. and Tian, Y., "A comprehensive survey of clustering algorithms", ***Annals of Data Science***, Vol. 2, No. 2, (2015), 165-193.
7. Hamidi, H. and Daraei, A., "Analysis of pre-processing and post-processing methods and using data mining to diagnose heart diseases", ***International Journal of Engineering-Transactions A: Basics***, Vol. 29, No. 7, (2016), 921-930.
8. Baron-Epel, O., Heymann, A.D., Friedman, N. and Kaplan, G., "Development of an unsupportive social interaction scale for patients with diabetes", ***Patient Preference and Adherence***, Vol. 9, (2015), 1033-1040.
9. Tomar, D. and Agarwal, S., "A survey on data mining approaches for healthcare", ***International Journal of Bio-Science and Bio-Technology***, Vol. 5, No. 5, (2013), 241-266.
10. Blangero, J., Teslovich, T.M., Sim, X., Almeida, M.A., Jun, G., Dyer, T.D., Johnson, M., Peralta, J.M., Manning, A. and Wood, A.R., "Omics-squared: Human genomic, transcriptomic and phenotypic data for genetic analysis workshop 19", in BMC proceedings, BioMed Central. Vol. 10, (2016), 20-29.
11. Farwell, K.D., Shahmirzadi, L., El-Khechen, D., Powis, Z., Chao, E.C., Davis, B.T., Baxter, R.M., Zeng, W., Mroske, C. and Parra, M.C., "Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: Results from 500 unselected families with undiagnosed genetic conditions", ***Genetics in Medicine***, Vol. 17, No. 7, (2015), 578-585.
12. Go, A.S., Mozaffarian, D., Roger, V.L., Benjamin, E.J., Berry, J.D., Blaha, M.J., Dai, S., Ford, E.S., Fox, C.S. and Franco, S., "Heart disease and stroke statistics—2014 update: A report from the american heart association", ***Circulation***, Vol. 129, No. 3, (2014), e28.
13. Li, H., Yimin, Z., Mengshi, C., Xun, L., Xiulan, L. and Ying LIANG, H.T., "Development and validation of a disease severity scoring model for pediatric sepsis", ***Iranian Journal of Public Health***, Vol. 45, No. 7, (2016), 875-883.
14. Stork, M. and Vancura, V., "Electronic evaluating system for pacemaker pulses optimization", in Applied Electronics (AE), International Conference on, IEEE., (2013), 1-4.
15. Miller, T., "Structure and physiology of the circulatory system", ***Comprehensive Insect Physiology, Biochemistry and Pharmacology***, Vol. 3, (2013), 289-353.
16. Gharagozlou, F., Saraji, G.N., Mazloumi, A., Nahvi, A., Nasrabadi, A.M., Foroushani, A.R., Kheradmand, A.A., Ashouri, M. and Samavati, M., "Detecting driver mental fatigue based on eeg alpha power changes during simulated driving", ***Iranian Journal of Public Health***, Vol. 44, No. 12, (2015), 1693.
17. Shadloo, B., Motevalian, A., Rahimi-Movaghar, V., Amin-Esmaeili, M., Sharifi, V., Hajebi, A., Radgoodarzi, R., Hefazi, M. and Rahimi-Movaghar, A., "Psychiatric disorders are associated with an increased risk of injuries: Data from the iranian mental health survey (iranmhs)", ***Iranian Journal of Public Health***, Vol. 45, No. 5, (2016), 623-630.
18. Moradi, S. and Hamidi, H., "Analysis of consideration of security parameters by vendors on trust and customer satisfaction in e-commerce", ***Journal of Global Information Management***, Vol. 25, No. 4, (2017), 32-45.
19. Hamidi, H., "A combined fuzzy method for evaluating criteria in enterprise resource planning implementation", ***International Journal of Intelligent Information Technologies (IJIIT)***, Vol. 12, No. 2, (2016), 25-52.
20. Hamidi, H., "A model for impact of organizational project benefits management and its impact on end user", ***Journal of Organizational and End User Computing (JOEUC)***, Vol. 29, No. 1, (2017), 51-65.
21. Johnson, R.D., Li, Y. and Dulebohn, J.H., "Unsuccessful performance and future computer self-efficacy estimations: Attributions and generalization to other software applications", ***Journal of Organizational and End User Computing (JOEUC)***, Vol. 28, No. 1, (2016), 1-14.
22. Kakar, A.S., "A user-centric typology of information system requirements", ***Journal of Organizational and End User Computing (JOEUC)***, Vol. 28, No. 1, (2016), 32-55.
23. Liu, Y., Tan, C.-H. and Sutanto, J., "Selective attention to commercial information displays in globally available mobile application", ***Journal of Global Information Management (JGIM)***, Vol. 24, No. 2, (2016), 18-38.

24. Mohammadi, K. and Hamidi, H., "Modeling and evaluation of fault tolerant mobile agents in distributed systems", in Wireless and Optical Communications Networks. WOCN. Second IFIP International Conference on, IEEE., (2005), 323-327.

25. Hamidi, H., Vafaei, A. and Monadjemi, S.A.H., "Analysis and evaluation of a new algorithm based fault tolerance for computing systems", ***International Journal of Grid and High Performance Computing (IJGHPC)***, Vol. 4, No. 1, (2012), 37-51.

26. Hamidi, H., Vafaei, A. and Monadjemi, S.A., "Analysis and design of an abft and parity-checking technique in high performance computing systems", ***Journal of Circuits, Systems, and Computers***, Vol. 21, No. 03, (2012), 1250017.

27. Hamidi, H. and Vafaei, A., "Evaluation of fault tolerant mobile agents in distributed systems", ***International Journal of Intelligent Information Technologies (IJIIT)***, Vol. 5, No. 1, (2009), 43-60.

28. Hamidi, H., Vafaei, A. and Monadjemi, S.A., "Evaluation and check pointing of fault tolerant mobile agents execution in distributed systems", ***Journal of Networks***, Vol. 5, No. 7, (2010), 800-807.

29. Hamidi, H., Vafaei, A. and Monadjemi, A., "A framework for fault tolerance techniques in the analysis and evaluation of computing systems", ***International Journal of Innovative Computing, Information and Control***, Vol. 8, No. 7, (2012), 5083-5094.

30. Wu, J., Ding, F., Xu, M., Mo, Z. and Jin, A., "Investigating the determinants of decision-making on adoption of public cloud computing in e-government", ***Journal of Global Information Management (JGIM)***, Vol. 24, No. 3, (2016), 71-89.

31. Chevers, D., Mills, A.M., Duggan, E. and Moore, S., "An evaluation of software development practices among small firms in developing countries: A test of a simplified software process improvement model", ***Journal of Global Information Management (JGIM)***, Vol. 24, No. 3, (2016), 45-70.

32. Bimonte, S., Sautot, L., Journaux, L. and Faivre, B., "Multidimensional model design using data mining: A rapid prototyping methodology", ***International Journal of Data Warehousing and Mining (IJDWM)***, Vol. 13, No. 1, (2017), 1-35.

33. Esposito, C. and Ficco, M., "Recent developments on security and reliability in large-scale data processing with mapreduce", ***International Journal of Data Warehousing and Mining (IJDWM)***, Vol. 12, No. 1, (2016), 49-68.

34. Hamidi, H. and Kamankesh, A., "An approach to intelligent traffic management system using a multi-agent system", ***International Journal of Intelligent Transportation Systems Research***, (2017), 1-13.

35. DARAEI, A. and HAMIDI, H., "An efficient predictive model for myocardial infarction using cost-sensitive j48 model", ***Iranian journal of public health***, Vol. 46, No. 5, (2017), 682.

36. Nilchi, A.N., Vafaei, A. and Hamidi, H., "Evaluation of security and fault tolerance in mobile agents", in Wireless and Optical Communications Networks,. WOCN'08. 5th IFIP International Conference on, IEEE. , (2008), 1-5.

37. Hamidi, H. and Mohammadi, K., "Modeling fault tolerant and secure mobile agent execution in distributed systems", ***International Journal of Intelligent Information Technologies (IJIIT)***, Vol. 2, No. 1, (2006), 21-36.

38. Hamidi, H., Vafaei, A. and Monadjemi, S.A., "A framework for abft techniques in the design of fault-tolerant computing systems", ***EURASIP Journal on Advances in Signal Processing***, Vol. 2011, No. 1, (2011), 90-91.

39. Abadi, A.G.R. and Hamidi, H., "Constrained model predictive control of low-power industrial gas turbine", ***International Journal of Engineering-Transactions B: Applications***, Vol. 30, No. 2, (2017), 207-214.

40. Hamidi, H. and Valizadeh, A., "Improvement of navigation accuracy using tightly coupled kalman filter", ***International Journal of Engineering-Transactions B: Applications***, Vol. 30, No. 2, (2017), 215-222.

41. Abadpour, M. and Hamidi, H., "Stabilization of v94. 2 gas turbine using intelligent fuzzy controller optimized by the genetic algorithm", ***International Journal of Applied and Computational Mathematics***, (2016), 1-14.

An Efficient Predictive Model for Probability of Genetic Diseases Transmission

H. Hamidi[a], F. Qaribpour[b]

Department of Industrial Engineering, Information Technology Group, K. N.Toosi University of Technology, Tehran, Iran

PAPER INFO

Paper history:
Received 05 May 2017
Received in revised form 18 June 2017
Accepted 07 July 2017

Keywords:
Genetic Disease
Transmitting Genetic Diseases
K-means Clustering Algorithm
Data Mining
Decision Trees

چکیده

در این مقاله، به تحلیل و ارزیابی برترین الگوریتم‌های استفاده شده در علم پزشکی برای پیش‌بینی انتقال بیماری‌های ژنتیکی پرداخته شده، و یک روش جدید ترکیبی از درخت تصمیم و خوشه‌بندی ارائه می‌شود. در این پژوهش تحت شرایط یکسان بر روی مجموعه داده‌ی استاندارد اعمال شده و بر اساس یکسری معیارهای اندازه‌گیری مانند مقدار پیش‌بینی مثبت، مقدار پیش‌بینی منفی، دقت، حساسیت و ویژگی، به مقایسه‌ی این الگوریتم‌ها برای پیش‌بینی دقیق‌تر انتقال بیماری پرداخته شده‌است. نتایج نشان می‌دهند که الگوریتم ماشین بردار پشتیبان عملکرد بهتری نسبت به دو الگوریتم ساده‌ی دیگر دارد و سپس شبکه عصبی و در آخر الگوریتم ژنتیک پیش‌بینی بهتری انجام داده‌است، درحالی‌که روش پیشنهادی ترکیبی با پارامترهای متفاوت ایجاد شده است؛ در مقایسه با روش‌های ساده عملکرد بهتری دارد.

doi*: 10.5829/ije.2017.30.08b.06*