A Hybrid Approach for Diabetes Risks Diagnosis Using Data mining and Neuro-Fuzzy System

Mr.C.Natarajan M.E.,(Ph.D)*
Dr. J.M. Gnanasekar M.E., Ph.D**

Abstract

The popularity of Data mining is ever-increasing in different research arenas due to its never-ending applications and methodologies to extract the information in proper manner .Health care industry produces huge quantity of data that control complex information relating to patients and their health conditions. Data mining techniques have the capabilities to determine unknown patterns or relations among the objects in the medical data. There is a lengthy record of diabetic registries and databases with systematically composed patient information. This Paper scrutinize one diabetic patients behavior data set, performing a method of classification by decision Tree Induction and diagnosing and reducing the diabetes risks by using Neural Networks and Fuzzy logic system. Data mining can facilitate healthcare organizations to expect trends in the patient conditions and their behaviors. Healthcare data mining provides innumerable possibilities for hidden pattern examination from these data sets. These pattern can be used to establish diagnoses, reduces the risks for the patients in healthcare organizations. The proposed Approach first deals with decision tree induction method to diagnose the risks of the diabetes patients whose ages are in between 30 and 60.Then outline the neural network and reduce the risks by adjusting the weights of the parameters and illustrate the results using the fuzzy logic. Every individual data mining Techniques provides different immense health care applications. While combining one or more data mining techniques, can get ideal skilled data and precise results .This research shows that, pooled techniques of mining such as Decision tree, neural networks and fuzzy logic approaches assist to afford enhanced and perfect treatment for reducing the risks of the diabetes patients.

Author correspondence:
C.Natarajan,
Ph.D Scholor, Department of Computer Science & Engineering
Saveetha University, Chennai, Tamilnadu, India
Corresponding Author Email Id: vcnataraj@gmail.com

* Ph.D Scholor, Deparment of Computer Science and Engineering, Saveetha University, Chennai, India
**Professor, Department of Computer Science and Engineering, Srivenkateswara college of Engineering, Chennai, India

1. Introduction

Diabetes might be one of the most dangerous diseases across the world and especially in India and USA,

but awareness about the same can well be estimated by the fact that India today has more people with type-2 diabetes. The World Health Organization also estimates that 80 per cent of diabetes deaths occur in low and middle-income countries and projects that such deaths will double between 2016 and 2030 [2].The diabetes capital of the world with as many as 50 million people suffering from type-2 diabetes, India has a challenge to face. However, medical experts feel that timely detection and right management can go a long way in helping patients lead a normal life. With the country having the highest number of diabetic patients in the world, the sugar disease is posing an enormous health problem to our country today. Often known as the diabetes capital of the world, India has been witnessing an alarming rise in incidence of diabetes according to the International Journal of Diabetes in Developing Countries. According to a World Health Organization (WHO) fact sheet on diabetes, an estimated 3.4 million deaths are caused due to high blood sugar. It has been further estimated that the global burden of type-2 diabetes is expected to increase to 438 million by 2030 from 285 million people (recorded in 2010). Similarly, for India this increase is estimated to be 58%, from 51 million people in 2010 to 87 million in 2030[1]. But debates, discussions and deliberations aside, the fundamental thing is to know what exactly is diabetes and how to prognosis and diagnosis diabetes diseases. Applying machine learning and data mining methods in this research is a key approach to utilizing large volumes of available diabetes-related data for extracting knowledge. In this framework, efforts were made to diagnose and reduce the risks of the diabetes.

Data mining tools can be very useful to control limitations of people such as bias or error due to fatigue, and to provide indications for the decision-making processes. The fundamental nature of data mining is in the identification of relations, patterns and models that provide support for predictions and of decision making process for diagnoses and treatment planning [3]. These models can be called predictive, and they are being integrated into information systems of hospitals as a model for decision making, reducing the risks and decision making time. Widespread use of information technology enables the elimination of manual tasks of data extraction from charts or filling of specialized questionnaires, extraction of data directly from electronic records, transfer on secure electronic system of medical records that will save lives and reduce the cost of health care, early detection of infectious diseases with advanced collection of data .Retrieval of information with the help of computers can help the quality of decision making and avoiding human errors. When there is a large volume of data that needs to be classified, decision making by people is usually poor [4] .Data mining represents the process of analyzing raw data with the help of computer and extraction of their meaning. It is frequently defined as discovering previously unknown and potentially useful information from large volume of data and it is possible to predict trends and customer behavior and thus provide the better health care treatment. This is accomplished by data analysis from various perspectives and finding the connections and relations between mutually unconnected information. In the process of data mining previously unknown trends and patterns from a database of historical information are being discovered and that information is being converted into useful patterns The following sections discuss about the related works that made by the other references, Research work explains the methodologies used in this research, Results and Discussions, Conclusions.

This section analytically and systematically reviews the literature on healthcare industry and categorizes the applications with data mining. A literature survey reveals many results on diabetes; the diabetic data warehouse was formed by a large integrated health care system in

the New Orleans area with more than 30,000 diabetic patients. Data mining has an infinite prospective to utilize healthcare data more resourcefully to predict different kind of disease (sheenal patel et al., 2016). A survey was made to know how to detect the contextual anomalies and hot spotting algorithm (Jianying Hu et al.,2012) The study was made to know how to apply data mining techniques to diagnose the diabetes for young and old patients(Abdulla A et al.,2012), the treatments for the risky patients can be identified. Another approach was recommended to demonstrate the importance of modifying risk factors for the development of diabetes and reducing the prevalence of diabetes in KSA (Al-Nozha et al.,2004). Healthcare institutions that use data mining applications have the possibility to predict future requests, needs, desires, and conditions of the patients and to make adequate and optimal decisions about their treatments. Prediction and decision making in healthcare domain by using datamining techniques studied from the literature review (Boris Milovic et al.,2012).Data mining indicated that education level did not predict changes in HbA1c levels (Sigur- dardottir et al., 2007). The status of insulin therapy for preschool age children with type1 diabetes has also been studied. In this study, conducted with data acquired over a 10 year period (1993–2002), the daily insulin therapy and episodes of severe hypoglycemia were identified in a population of approximately 150 patients diagnosed with diabetes at less than 6 years of age (Yokotaa et al., 2005).

Datamining Approach for prediction and treatments for diabetic disease study was conducted at International journal of Science Innovations today (velidePhani kumar et al., 2014). Hyunjung Shin et al. (2012) state that mining hospital records, such safety issues could be flagged and addressed by hospital management and government regulators.A study was made to Extracting cardiac dynamics within ECG signal for human identification and cardiovascular diseases classification using neural networks (MuqingDeng et al.,2018),cardiac dynamics within ECG signals is extracted (approximated) accurately by using radial basis function (RBF) neural networks through deterministic learning mechanism. The obtained cardiac system dynamics is represented and stored in constant RBF networks.

2. Research Method

Hybrid System Framework is a new approach that combines three datamining techniques to produce the better result. First the framework starts with decision tree classification method used to make the decision of the patient risks based up on their behaviors, and then the risks levels are given as the input layer parameters to the Neural Network layered architecture. Here by using perceptron training algorithm, adjusts the weights of the parameters, minimize the risks levels to nil. Fuzzy logic systems draw the evaluation graphs for the medium, high and low risks levels.

Research design:

Decision tree models are best suited for data mining. They are inexpensive to construct, easy to interpret, easy to integrate with database system and they have comparable or better accuracy in many applications[7]. The data set contains three predictor attributes, namely Age, Gender, Alcohol Consumptions and one goal attribute, risks whose values indicates the risks level of the patients. Decision tree shown in figure 1.a can be used to classify an unknown class data instance with the help of the data set given. The idea is to push the instance down the tree, following the branches whose attributes values match the instances attribute values, until the instance reaches a leaf node [8].

For example, the data instance to be classified is described by the tuple (Age=23, Gender=Female, Alcohol = yes, Risk =?), where "?" denotes the unknown value of the goal instance. The tree tests the Age value in the instance. If the answer is 23(between 30 and 45),

the instance is pushed down through the corresponding branch. The tree tests the Alcohol value in the instance. If the answer is medium; the instance is pushed down through the corresponding branch. Now the instance reaches the leaf node, where it is classified as High. Neural Network layered architecture and fuzzy logic systems are shown the above figure 1.b and figure 1.c respectively.

Research Procedure:

This proposed approach combines the different methodology of data mining to produce the better result. The hybrid methodologies are Decision Tree Classification, Neural Networks and Fuzzy logic system.

Decision tree:

The knowledge representation structure consisting of nodes and branches prepared in the form of a tree such that, every internal non-leaf node is labeled with name of the attributes. The branches coming out from an internal node are labeled with values of the attributes in that node. Every node is labeled with a risk class. Decision tree models are best suited for data mining [13]. They are inexpensive to construct, easy to interpret, easy to integrate with database system and they have comparable or better accuracy in many applications. Experiments are conducted by using the training data set of 900 instances with 11 different attributes. Depending upon the attributes, the dataset is classified into two parts, i.e. 70% of the data is used for training and 30% is used for testing. Data set and the parameters lists are consider by the behavior of the patients.

Decision tree algorithm:

Key Parameters: Age, gender, marital status, Alcohol Consumption, Blood pressure, Plasma, Insulin level, type of diabetes, Nature of job, Exercise, Diet Control.

Weights of an parameters are in the range of {0.5, 1, 1.5}

Age [] = {31-40, 41-50, 51-60};

Marital Status [] = {Unmarried, Divorced, Married};

Alcohol Consumptions [] = {nil, rare, regular};

Blood pressure [] = {low (<90), medium (100-140), high (>140)};

Nature of job [] = mind tension and thinking time {low, medium, high};

Exercise [] = {regular, rare, nil};

Diet Control [] = {good, ok, bad};

All the tuples having all 10 parameters but following 6 are vital parameters

Tuple{marital ,alcohol,B.p,job,ex,diet}= t1{1.5,1,1,1,1.5,1},tot=7;

If (tot is < 3) then risks = low;

Else if (tot is 3 to 6) then risks = medium;

Else risks = high risky;

Neural Network:

In this hybrid approach, neural networks are used to reduce the risks level obtained from the decision tree classification. By adjusting the weights of the parameters the risks level of the diabetes can be reduced to nil. Neural network is a mathematical model or computational model based on biological neural network[15]. Neural networks process information in a similar way the human brain does. The network is composed of a large number of highly interconnected processing neurons working in parallel to solve a specific problem [1].

Fuzzy logic system:

This proposed approach uses the fuzzy set theory to note down the risks level of the diabetes.The sets can be distinguished by high, medium, low.Fuzzy set theory and fuzzy logic are highly suitable for developing knowledge based systems in healthcare for diagnosis of risk

of diseases. For this, sugeno model of fuzzy system [20] is used. The fuzzy rules are generated based on experts knowledge in this domain, only 6 attributes are found to be effectual and required for diabetes prediction. In the proposed system, the input is the set of all the selected features and the output of the system is to achieve a value 0 or 1 that indicates the nil risks and high risks of diabetes in patients.The following data sets shown the sample analysis for the risks calculation.
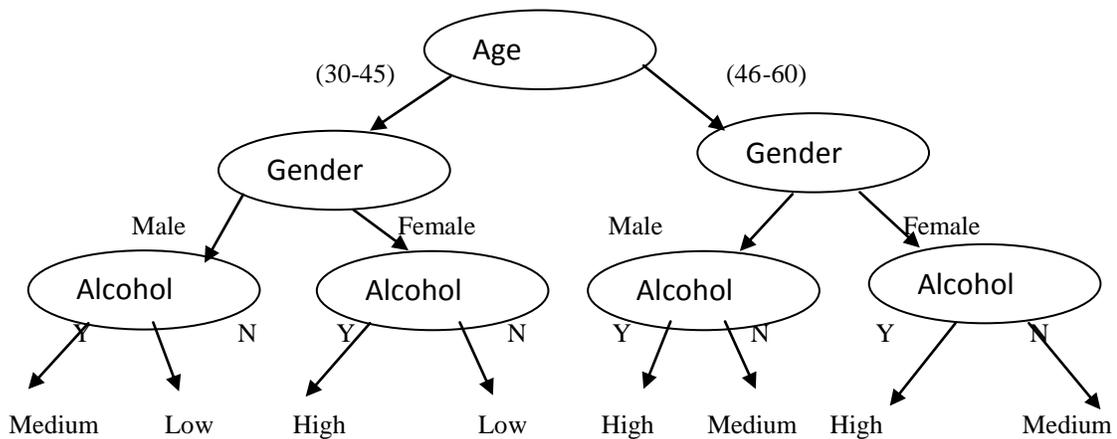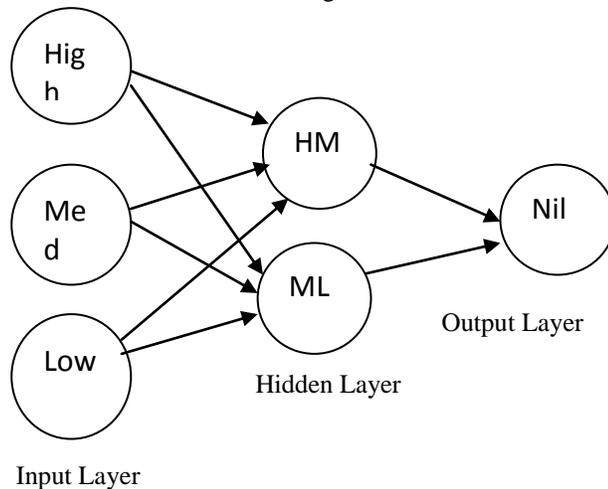

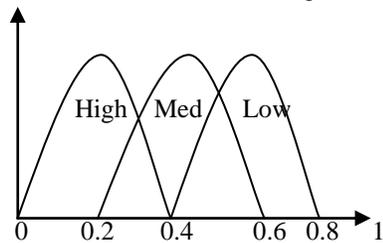Figure 1.a: Decision Tree Classification

Figure 1.b: Neural Network Layers

Figure 1.c : Fuzzy Logic System

Figure 1 :Hybrid System Framework

Table 1: Sample data set

Marital status	Alcohol	B.p	Job	Exercise	Diet	Diabetes Risks probability
M	nil	high	low	nil	bad	High
UM	rare	low	med	rare	Ok	Medium
M	nil	med	low	regular	good	low
M	nil	Low	med	nil	good	Medium
M	nil	low	low	regular	Ok	low
M	regular	high	med	nil	Ok	High
M	regular	med	med	nil	good	High
D	regular	high	med	rare	good	High
M	rare	med	med	regular	ok	Medium
UM	regular	Low	high	nil	bad	High

3. Results and Analysis

The risks reduction accuracy of this hybrid approach includes the combination of Decision Tree, neuro-fuzzy systems is around 95%. These comparison results of other different data mining applications for reducing the diabetes risks and the success rate of these applications are shown in Table 2 and

Fig. 2.This Analysis clearly shows that the Hybrid Approach of Decision Tree, Neural networks and fuzzy logic system gives the better performance than those Techniques applied separately.

Table 2: Comparison of Hybrid system Performance over Decision Tree, Neural networks

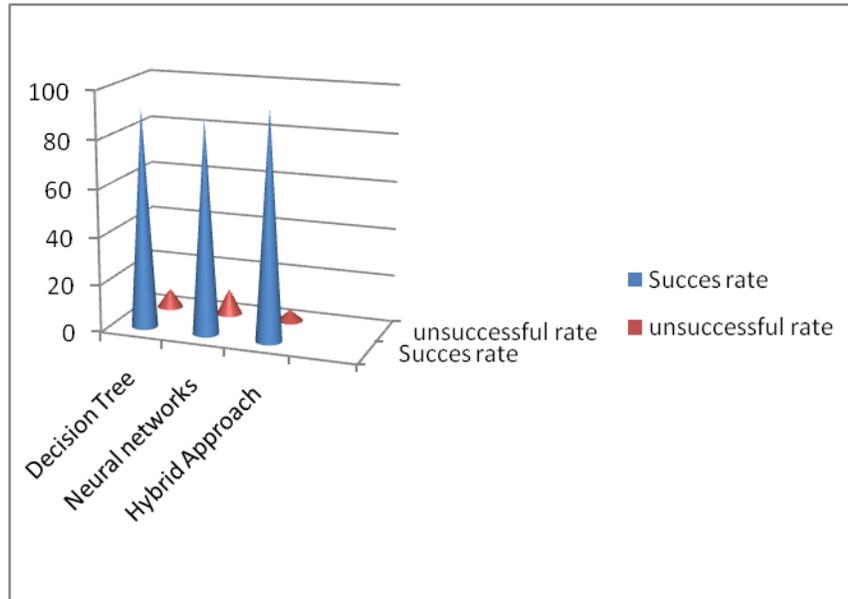| Data mining Techniques	% of Success rate	% of Unsuccessful rate
Decision Tree	92.01	7.99
Neural networks	89.23	10.77
Hybrid System (Decision tree,NN)	95.4	4.6

Figure 2: Graphical Analysis of Hybrid System with other techniques

4. Conclusion

This paper combined the different data mining application in the healthcare sector for diagnose the risks of the diabetes diseases and make the way to reduces it.The prediction of diseases risks and reducing using Data Mining applications is a challenging task but it drastically reduces the human effort and increases the diagnostic accuracy. Mounting an efficient data mining tools for an application could reduce the cost and time constraint in terms of human resources and expertise. The comparison study shows the interesting results that data mining techniques in all the health care applications give a more encouraging level of accuracy 95.4% for diabetes risks diagnoses and reduction. The future system will give the maximum percentage accuracy and plan to deploy the web applications and make useful to all the web users.

References

[1]   Ioannis Kavakiotis, Olga Tsave, Athanasios Salifoglou, Nicos Maglaveras,b, Ioannis Vlahavas, and Ioanna Chouvarda "Machine Learning and Data Mining Methods in Diabetes Research ",*Comput Struct Biotechnol J. 2017; 15: 104–116. PMCID: PMC5257026* Published online 2017 Jan 8. doi: 10.1016/j.csbj.2016.12.005

[2]   Rakesh Malik.," India is the diabetes capital of the world", *Times of India | Mumbai Mirror |* Jan 28, 2016, 12:41 IST

[3]   Candelieri, A., Dolce, G., Riganello, F., & Sannita, W. G. (2011). *Data Mining in Neurology. In Knowledge- Oriented Applications in Data Mining* (pp. 261-276). InTech.

[4]   Bushinak, H., AbdelGaber, S., & AlSharif, F. K. (2011). "Recognizing The Electronic Medical Record Data From Unstructured Medical Data Using Visual Text Mining TechniquesProf. Hussain Bushinak", *(IJCSIS) International Journal of Computer Science and Information Security,* Vol. 9, No. 6 , 25-35.

[5]   Eapen, A. G. (2004).,"Application of Data mining in Medical Applications".*Ontario, Canada, 2004: University of Waterloo.*

[6]   Milovic, B. (2011). "Usage of Data Mining in Making Business Decision". *YU Info 2012 & ICIST 2012,* (pp. 153-157).

[7]   boirefillergroup.com. (2010). "Data Mining Methodology". *Retrieved 06 12, 2012, from Boire-Filler Group: http://www.boirefillergroup.com/methodology.php*

[8]   Velidephani Kumar-Et Al., "A Data Mining Approach For Prediction And Treatment Of diabetes Disease" *Ijsit, 2014*, 3(1),073-079

[9]   Abdullah A. Aljumah, Mohammed Gulam Ahamad, "Application of data mining: Diabetes health care in young and old patients*", Journal of King Saud University – Computer and Information Sciences (2013)* 25,127-136

[10]   Boris Milovik,Milan Milovic, "Prediction and Decision making in Healthcare using Datamining", *General Hospital Vrbas, Serbia ,International journal of Public Health Science Vol 1, No.2,December 2012*,pp 69-78,ISSN:2252-8806

[11]   Zaïane, O. R. (1999). "Principles of Knowledge Discovery in Databases", *Department of Computing Science, University of Alberta.*

[12]   Gupta, S., Kumar, D., & Sharma, A. (2011). "Data Mining Classification Techniques Applied For Breast Cancer Diagnosis And Prognosis",. *Indian Journal of Computer Science and Engineering (IJCSE) 188-195.*

[13]   Khan, F. S., Anwer, R. M., Torgersson, O., & Falkman, G. (2008). "Data Mining in Oral Medicine Using Decision Trees". *World Academy of Science, Engineering and Technology 37,* (pp. 225-230).

[14]   Ngan, P. S., Wong, M. L., Lam, W., Leung, K. S., & Cheng, J. C. (1999). "Medical data mining using evolutionary computation". *Artificial Intelligence in Medicine 16, (pp. 73–96).*

[15]   Shouman, M., Turner, T., & Stocker, R. (2012). "Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients. 2012",*International Conference on Knowledge Discovery (ICKD 2012) IPCSIT Vol. XX. Singapore: IACSIT Press.*

[16]   Stühlinger, W., Hogl, O., Stoyan, H., & Müller, M. (2000). "Intelligent Data Mining for Medical Quality Management. Proc", *Fifth Workshop Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP2000), Workshop Notes of the 14th European Conf. Artificial Intelligence.*

[17]   Koyuncugil, A. S., & Ozgulbas, N. (2010). "Donor Research and Matching System Based on Data Mining in Organ Transplantation.", *J Med Syst (2010) 34 ,* 251–259.

[18]   Dhanya P Varghese & Tintu P B, "A Survey On Health Data Using Data Mining Techniques", *International Research Journal of Engineering and Technology (IRJET), Volume:02 Issue: 07*, Oct-2015

[19]   W.L.Zuoa & Z.Y.Wanga & T.Liua & H.L.Chenc, "Effective Detection Of Parkinson's Disease Using An Adaptive Fuzzy K-Nearest Neighbor Approach", *Biomedical Signal Processing and Control, Elsevier,* pp. 364373, 2013