International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Intelligent Prediction of Heart Disease Diagnosis Using ANFIS Classification Model

Shraddha. R. Warhade[1], U. W. Hore[2]

[1]P. G. Scholar, EXTC Dept. P.R. Pote (Patil) College of Engineering and Management, Amravati, Maharashtra, 444601, India

[2] Professor, EXTC Dept. P.R.Patil Collage of Engineering & Technology, Amravati, Maharashtra, 444601, India

Abstract: ***Cardiovascular diseases are being to become the main cause of death in most countries of the world. The considerable growing of cardiovascular disease and its effects and complications as well as the high costs on society makes medical community seek for solutions to prevention, early identification and effective treatment with lower costs. Thus, valuable knowledge can be established by using artificial intelligence , The discovered knowledge makes improve the quality of service. Heart disease is a term that assigns to a large number of medical conditions related to heart. These medical conditions describe the abnormal health conditions that directly influence the heart and all its parts. Heart disease is a major health problem in today's time. Diagnosis of heart disease by using machine learning methods is one of the challenges in the health field. Technically, the ANFIS performs a vital role for prediction of diseases in medical industry . ANFIS is a fuzzy inference system (FIS) implemented in the framework of an adaptive fuzzy neural network. It combines the explicit knowledge representation of an FIS with the learning power of artificial neural networks. The objective of ANFIS is to integrate the best features of fuzzy systems and neural networks. This paper will be provided on a particular dataset using classification and feature selection approach. In this we will use feature ranking on effective factors of disease related to Cleveland clinic database and by using Novel Feature Selection Method as well as ANFIS, 13 effective factors reduced to 5 optimized features in terms of accuracy. The assessment of selected features of classified methods also showed that NFS method along with ANFIS has the best accurate criteria of the rate of 97.61% on these features.***

Keywords: Heart Disease Diagnosis, Novel Feature Selection Method , Adaptive Neuro Fuzzy Inference System (ANFIS), Classification, Feature Selection.

1. Introduction

Life is dependent on efficient working of heart because heart is essential part of our body. If operation of heart is not proper, it will affect the other body parts of human such as brain, kidney etc. Heart disease is a disease that affects on the operation of heart. Heart and blood vessel disease also called heart disease includes numerous problems, many of which are related to a process called atherosclerosis. Atherosclerosis is a condition that develops when a substance called plaque builds up in the walls of the arteries. This buildup narrows the arteries, making it harder for blood to flow through. If a blood clot forms, it can stop the blood flow. This can cause a heart attack or stroke. There are number of factors which increases risk of Heart disease. High blood cholesterol and high triglyceride levels ,high blood pressure ,diabetes and pre diabetes , overweight and obesity , smoking, lack of physical activity , unhealthy diet, stress .

Nowadays, in the world, Heart disease is the major cause of deaths which is also supported by the World Health Organization survey which has estimated that 12 million deaths occur worldwide, every year due to the Heart diseases.. As reported by America Heart Association, about 525,000 people suffer an early heart attack and about 210,000 people suffer repeatable ones, each year [2]. According to the reports of Center for Disease Control (CDC), almost more than half of the deaths in 1999, occurred before any emergency services and hospital treatments to be done on the patient [3]. Cardiovascular disease is a condition of the heart in which the patient may suffer a heart attack, angina, heart failure and so on. In the past, epidemics and infectious diseases were a major cause of human death, but today noncommunicable diseases have become more dangerous factors. Figure (1) shows ten main factors have led to the deaths over the past decade [4]. It is also the chief reason of deaths in numerous developing countries. Heart disease kills one person every 34 seconds in the United States.

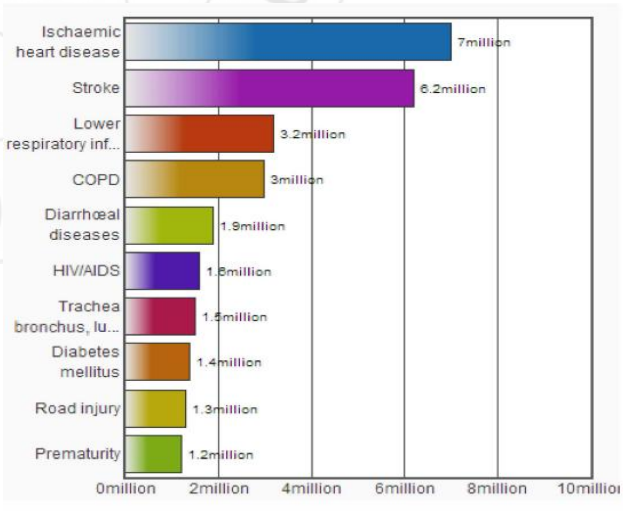

Figure 1: Ten leading causes of death in the world over the past decade

Eminent health services imply diagnosing patients correctly and providing effective treatment. Because a silent heart attack does not produce symptoms that cause the victim to seek medical help, the diagnosis is only made after the fact—after the damage has been done Inferior medical decisions can lead to disastrous consequences which are not

Licensed Under Creative Commons Attribution CC BY

acceptable. Health industry must also try to minimize the number of tests for identifying the disease. All this can be achieved by commissioning appropriate decision support system. Today many healthcare organizations have employed hospital information systems for managing patient's data. Unfortunately this data is not used for decision making. This huge data can be used for answering questions like "predict the probability of patients getting heart disease". Appropriate computer-based information and/or decision support systems can aid in achieving clinical tests at a reduced cost. The purpose of this research is to develop a model for prediction based on the demographic data of the patients' and provide best accurate result using ANFIS (adaptive neuro fuzzy interference system).

As purpose , the main aim of this research is to develop the system which improve the heart disease diagnosis using ANFIS classification model. Technically, the ANFIS performs a vital role for prediction of diseases in medical industry. Diagnosis of heart disease by using machine learning methods is one of the challenges in the health field. This research will be provided on a particular dataset using classification and feature selection approach. In this we will use feature ranking on effective factors of disease related to Cleveland clinic database and by using Novel Feature Selection Method as well as ANFIS, 13 effective factors reduced to 5 optimized features in terms of accuracy**.**

2. Literature Review

In this paper we survey different papers in which one or more method used for prediction of heart disease. The related work consists of attributes containing patient's medical history and symptoms. Majid Ghonji Feshki, Omid Sojoodi Shijani [5] proposed the improving the heart disease diagnosis by evolutionary algorithm of pso and feed forward neural network. In this research the work has been provided on a particular dataset using classification and feature selection approach.and provide the best accurate criteria for heart dieses diagnosis. The assessment of selected features of classified methods also showed that PSO method along with Neural Networks of Feed Forward Back-Propagation has the best accurate criteria.

Beant Kaur,Williamjeet Singh[6] proposed review on heart disease prediction system using data mining techniques . Data mining provides the methodology and technology to transform mounds of data into useful information for decision making and use this data for the prediction of heart disease with good accuracy. So that the prediction by using data mining algorithm given efficient results. Applying data mining techniques to heart disease treatment data can provide as reliable performance as that achieved in diagnosing heart disease.

Jyoti Soni ,Ujma Ansari, Dipesh Sharmah,Sunita Soni[7] proposed predictive data mining for medical diagnosis. the problem of constraining and summarizing different algorithms of data mining used in the field of medical prediction are discussed. The focus is on using different algorithms and combinations of several target attributes for intelligent and effective heart attack prediction using data mining. They conclude that the accuracy of the Decision Tree and Bayesian Classification further improves after applying genetic algorithm to reduce the actual data size to get the optimal subset of attribute sufficient for heart disease prediction.

K.Prasanna Lakshmi, Dr. C. R .K.Reddy [8] proposed fast rule-based heart disease prediction using associative classification mining. This research uses associative classification which builds a classifier with prediction rules of high interestingness values using Associative Classification and rewarding approach and experimental results show that this work helps doctors in their diagnosis decisions**.**

N. Aditya Sundar1, P. Pushpa Latha2, M. Rama Chandra3 [9] proposed performance analysis of classification data mining techniques over heart disease data base. This paper describes about a prototype using data mining techniques, namely Naïve Bayes and WAC (weighted associative classifier).They used two data mining classification modeling techniques and extracts hidden knowledge from a historical heart disease database. DMX query language and functions are used to build and access the models. . Data Mining Extension query language was used for model creation, model training, model prediction and model content access and Classification Matrix methods are used to evaluate the effectiveness of the models. The two models are able to extract patterns in response to the predictable state.

The most common methods used by researchers for diagnosis and monitoring of cardiovascular disease are classification and clustering methods. However, what has been less into the attention so far is the investigation of the issue by several types of methods and reaching a consensus in order to diagnose heart disease with highest precision and availability in the shortest time and the lowest cost as possible. Therefore, the purpose of the present study was to select the best features with the lowest costs and shortest times and highest precisions using ANFIS classification model

3. Proposed Work

The block diagram of prediction of heart disease diagnosis system using anfis is computer based automatic medical diagnosis system is illustrated in figure 2. Based on the principle of ANFIS, this research has five processes initiated by data collection, attribute and data selection, data initialization, modeling, and evaluation. The scope of method in this research has been shown

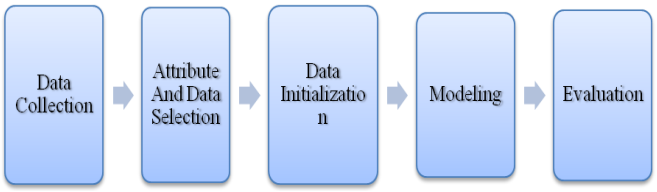

Figure 2: Block Diagram Of Prediction Of Heart Disease Diagnosis System Using ANFIS

Licensed Under Creative Commons Attribution CC BY

3.1 Data Collection

In health, the data is mostly extracted from the client files at health centers. To conduct this research the data related to Cleveland clinic existing in UCI machine learning repository were used. This dataset has collected the data related to 303 healthy and sick people from which 138 were sick and 165 were healthy. This databaset contains 76 attributes, but all publications refer to using a subset of 14 of them, which 13 are related to disease factors and 1 case is related to the presence or absence of heart disease in people who underwent 13 previous tests.

3.2 Attribute and Data Selection

It must be noted that these 14 factors are in fact the most important and most basic factors ever identified to diagnose heart attacks; therefore the investigation on these factors can be considered quite precise and important. Feature selection is the process of selecting a subset of relevant features to be used in constructing a classifier. It improve the diagnosis performance and provides a faster classifier. because of these advantages ,we execute a fourth trial to compare classification accuracy using all the 13 features with that of five best feature selected by novel feature selection algorithm [17] that are thalach, exang, oldpeak, slope, thal . The feature selection process can not only reduce the cost of recognition by reducing the number of features that need to be collected, but also improve the classification accuracy of the system.

3.3 Data initialization

The dataset is divided into two parts that is 40% of the data are used for training and 60% are used for testing .Using a given input/output data set, the toolbox function ANFIS train and initialize a fuzzy inference system (FIS) whose membership function parameters are tuned (adjusted) using either a backpropagation algorithm alone, or in combination with a least squares type of method. This adjustment allows your fuzzy systems to learn from the data they are modeling.

3.4 Modeling

The modeling approach used by ANFIS is similar to many system identification techniques. An adaptive neuro fuzzy inference system or adaptive network based fuzzy inference system (ANFIS) is a kind of artificial neural network that is based on Takagi–Sugeno fuzzy inference system. Since it integrates both neural networks and fuzzy logic principles, it has potential to capture the benefits of both in a single framework. Its inference system corresponds to a set of fuzzy IF–THEN rules that have learning capability to approximate nonlinear functions . Model validation is the process by which the input vectors from input/output data sets on which the FIS was not trained, are presented to the trained FIS model, to see how well the FIS model predicts the corresponding data set output values.

We use a validation data set to check and control the potential for the model overfitting the data. When checking data is presented to anfis as well as training data, the FIS model is selected to have parameters associated with the minimum checking data model error.

3.5 Evaluation

After the training data and generating the initial FIS structure, we can start testing the FIS and evaluate the result of the system that person is healthy or not .if not then which level of heart disease of that person. Evaluation is the process by which the input vectors from input/output data sets on which the FIS was not trained, are presented to the trained FIS model, to see how well the FIS model predicts the corresponding data set output values

4. Result Analysis

In our Implementation, the test of proposed technique to diagnose the heart disease is performed. We used UCI (Cleveland clinic) database of different patient which consist of 303 people data. These data are taken as input to the system .Also this database contains 76 attributes, but all publications refer to using a subset of 14 of them, which 13 are related to disease factors and 1 case is related to the presence or absence of heart disease in people but we use only 5 of them by NFS method which are important to heart disease factor. We use ANFIS model which load, train and test the data and finally evaluate the result.

4.1 Performance parameter

In our research to access the effectiveness of the system to diagnose the heart disease accuracy, recall, precision criteria, training and testing time consider .These criteria are calculated using the result of applying ANFIS method on data. Accuracy, recall and precision criteria was calculated from equation 1,2,3 and confusion matrix shown in fig 3.

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

$$\text{Recall} = \frac{TP}{(TP+FN)} \quad (2)$$

$$\text{Precision} = \frac{TP}{(TP+FP)} \quad (3)$$

4.2 Performance table

For create the confusion matrix we use 42 people data from dataset, from the matrix shown in figure 3, TP is the number of samples that are healthy and properly diagnosed are 25. TN means the number of samples that are sick and properly diagnosed are16. FP indicates the number of samples that are healthy and have been diagnosed wrongly is 0. FN indicates the number of samples that were sick but healthy wrongly diagnosed is 1.

Licensed Under Creative Commons Attribution CC BY

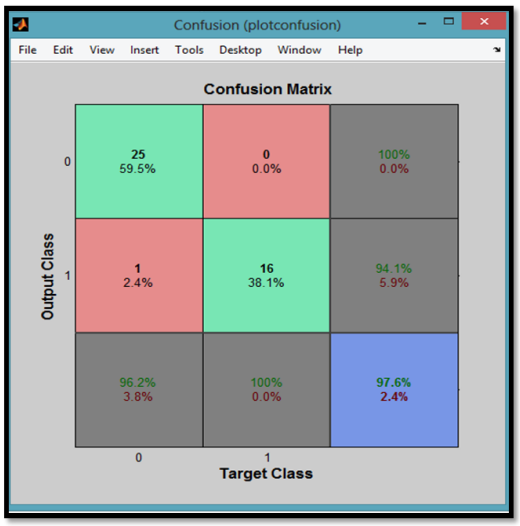

Figure 3: Plot Confusion Matrix

On the other hand, considering the feature selection and measurement of 5 features as input from 13, improve the time for diagnosis which result in improving the acceleration of test, accessibility and venture. Table 1 shows performance table which shows the accuracy, average testing time, precision and recall of our proposed work. Which is calculated from plot confusion matrix and performance parameter equation.

Table 1: Performance Table

Accuracy	Average testing time	Precision	Recall
97.61%	0.0107 sec	100%	96.15%

4.3 Comparison Table

In conducted investigations, many studies have been done to increase the accuracy of prediction of heart disease. As mentioned, this study has been evaluated using the specific data information from the Cleveland clinic. In this section the results of other studies on the same data are compared with available survey data. So far, the three approaches of neural networks, clustering and classification are used in the diagnosis of heart disease.

In this section the articles and researches with the UCI dataset activities and accuracy of each along with comparing those with results of this research are presented in Table 2

As shown in Table 2, from the various methods, proposed method (ANFIS with NFS, not only reduces the cost (13 features to 5 optimized features) but by far has had the greatest amount of scale accuracy. This approach has been able to increase accuracy standards by % 5.67 .

Table 2 : Comparison Table

Authors	Algorithm	Features	Accuracy	Feature Ranking
Panday and Godara, 2012	Improved MLP	13	82.80%	No
Subbalakshmi., 2011	Nalve Bayes	15	81%	No
Nahal, 2013	SMO	13	86%	No
Kumari And Godara 2011	SVM	13	84.12%	No
Srinivas., 2010	ODANB	13	83.70%	No
EI-Rashidy., 2010	Fuzzy Clustering	13	63.15%	No
Abdullah And Rajalaxmi, 2012	Decision Tree	13	63.33%	No
Soni., 2011	Association Rules	13	81.51%	No
Khemphila , Boonjing, 2011	Back-Propagation MLP	8	89.56%	Yes
Majid AND Omid, 2016	PSO+FFBP	8	91.94%	Yes
OUR WORK, 2017	ANFIS	5	97.61%	Yes

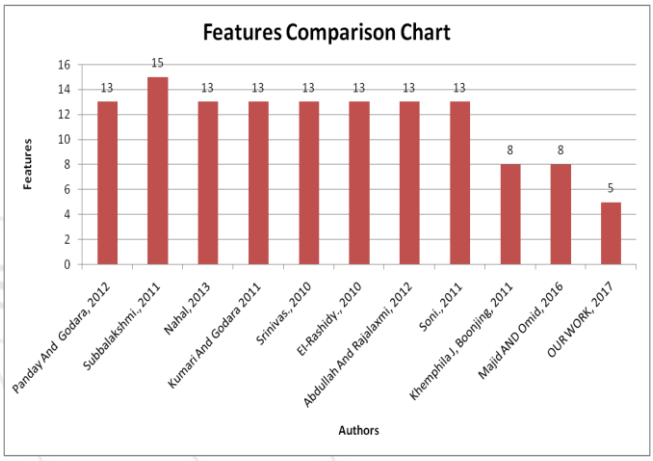

Figure 4: Feature Comparison Chart

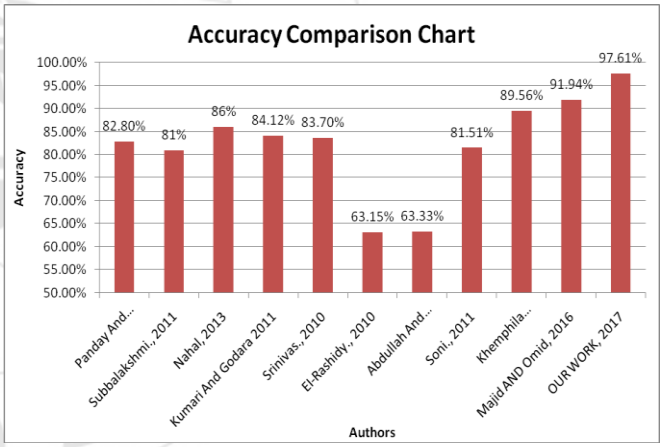

Figure 5: Accuracy Comparison Chart

In our experimentation we have even compared accuracy and features of our methodology with different researchers and there method used to diagnose the heart disease shown in figure 4 and 5 we provide maximum accuracy using minimum features than other methodology .

5. Conclusion and Future Scope

5.1 Conclusion

The considerable growing of cardiovascular disease and its effects and complications, diagnosis of heart disease by using machine learning methods is one of the challenges in the health field. In this study, diagnosis of heart disease based on artificial neuro fuzzy interface classification system is analyzed. Technically, the ANFIS performs a vital role for prediction of diseases in medical industry . In this we will use

Licensed Under Creative Commons Attribution CC BY

feature ranking on effective factors of disease related to Cleveland clinic database by using ANFIS, 13 effective factors reduced to 5 optimized features in terms of accuracy. The Novel Feature Selection method provides feature selection process to optimize the classification ability based on training data and to predict future cases. With a choice best subset includes 5 characteristics of Exercise tests (Slope, Oldpeak, and Exang), Thalach (heart rate), Thal accuracy have been improved by %5.67 and the accuracy has reached 97.61%. Also, the costs (including price, time, and complexity) have reduced due to considering feature selection.

The analysis also shows that different technologies are used in all the papers with taking different number of attributes. So, different technologies used shown the different accuracy to each other. But comparatively other papers , ANFIS provide best accuracy by using minimum number of attributes. This decision support system can use for providing better health services to heart patient. Thus, the early diagnosis of heart disease detection may decrease the chances of death in cardiac.

5.2 Future Scope

As a future work in this area, it is recommended that this method can be extended addition to feature selection (reduce the attributes) being based on accuracy and costs (in terms of accessibility, venture, testing time and price), the features are selected in such a way that the order of choosing features or in fact, the tests are mentioned as well

References

[1] Panday, P. and N. Godara, "Decision support system for cardiovascular heart disease diagnosis using improved multilayer perception." International Journal of Computer Applications, 2012. 45(8): p. 12-20
[2] Danaei, G., et aI., The preventable causes of death in the United States: comparative risk assessment of dietary, lifestyle, and metabolic risk factors. PLoS medicine, 2009. 6(4): p. e1000058.
[3] Control, C.f.D. and Prevention, State-specific mortality from sudden cardiac death--United States, 1999. MMWR. Morbidity and mortality weekly report, 2002. 51(6): p. 123.
[4] Available from :http://www. who.int lmediacentrel factsheets Ifs31O./
[5] Majid Ghonji Feshki, Omid Sojoodi Shijani, "the Improving the Heart Disease Diagnosis by Evolutionary Algorithm of PSO and Feed Forward Neural Network" ,IEEE 2016.
[6] Beant Kaur, Williamjeet Singh "Review on Heart Disease Prediction System using Data Mining Techniques." International Journal on Recent and Innovation Trends in Computing and Communication,2014
[7] Jyoti Soni ,Ujma Ansari, Dipesh Sharmah,Sunita Soni , "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction". International Journal of Computer Applications (0975 – 8887)Volume 17– No.8, March 2011
[8] K.Prasanna Lakshmi, Dr. C. R .K.Reddy , "Fast Rule-Based Heart Disease Prediction using Associative Classification Mining." IEEE International Conference on Computer, Communication and Control (IC4-2015).
[9] N. Aditya Sundar1, P. Pushpa Latha2, M. Rama Chandra3, " Performance analysis of classification data mining techniques over heart disease data base." International journal of engineering science & advanced technology
[10] Teerapat Kansadub1,Sotarat Thammaboosadee2, Supaporn Kiattisin3 " stroke risk prediction model based on demographic data." The 2015 Biomedical Engineering International Conference
[11] Subbalakshmi, G., K. Ramesh, and M.C. Rao, "Decision support in heart disease prediction system using naive bayes". Indian Journal of Computer Science and Engineering (IJCSE), 2011. 2(2): p .170-176.
[12] Khemphila, A. and V. Boonjing. Heart disease classification using neural network and feature selection. in Systems Engineering (ICSEng), 2011 21st International Conference on. 2011. IEEE.
[13] Abdullah, A. and R. Rajalaxmi. A data mining model for predicting the coronary heart disease using random forest classifier.
[14] Soni, .I., et aI., Intelligent and effective heart disease prediction system using weighted associative classifiers. International Journal on Computer Science and Engineering, 2011. 3(6): p. 2385-2392.
[15] Nahar, .I., et aI., Computational intelligence for heart disease diagnosis: A medical knowledge driven approach. Expert Systems with Applications, 2013. 40(1) :p. 96-104.
[16] Srinivas, K., B.K. Rani ,and A. Govrdhan, Applications of data mining techniques in healthcare and prediction of heart attacks. International Journal on Computer Science and Engineering (IJCSE), 2010. 2(02): p. 250-255.
[17] R.Suganya, S. Rajaram, A. Sheik Abdullah and V. Rajendran, "a novel feature selection method for Prediction heart diseases with data mining techniques" asian journal of information technology15(8):1314-1321

Author Profile

Miss. Shraddha R. Warhade received Bachelor of Engineering in Electronics and Telecomm. from SGB Amravati University and Pursuing Master of Engg. from P.R.Pote (Patil) COEM, Amravati.

Prof. Umesh W. Hore currently working as Asst. Professor , Department of Electronics & Tele-communications in P. R. Patil Collage Engineering and Technology. Amravati. His area of Specialization is microcontroller and microprocessor.

Licensed Under Creative Commons Attribution CC BY