Arrhythmia Disease Prediction System (ADPS) Using Machine Learning Concepts

B.Bavani [1], S. Nirmala Sugirtha Rajini [2], M.S. Josephine [3], V. Prasannakumari [4]

[1]Research Scholar, Department of Computer Applications, Dr. MGR. Educational and Research Institute, Chennai.
[2]Professor, Department of Computer Applications, Dr. MGR. Educational and Research Institute, Chennai.
[3]Professor, Department of Computer Applications, Dr. MGR. Educational and Research Institute, Chennai.
[4]Professor, Department of Information Technology, Rajalakshmi Engineering College, Thandalam, Chennai.

Abstract

Lifestyle changes refer to modifying things, so automatically the food habits is also changed. Due to this reason many people are affected by some dreadful diseases like blood pressure, cholesterol, and some heart related diseases. Among these the most dangerous disease is heart diseases, this is the main reason for increasing death rate. So there is a need to predict the heart disease in the earlier manner, the earlier prediction of the heart diseases is easy to cure. In medical domain the usage of machine learning concepts is growing every day. In this proposed system Machine Learning Algorithms are applied to forecast the Arrhythmia disease in earlier manner. Arrhythmia is one of the heart disease which means that the heart beats to quickly or too slowly or sometimes with an abnormal pattern. Here to detect the disease by using KNN algorithm, Naïve Bayes and decision tree. Finally the performance of these machine learning algorithms are analyzed in terms of accuracy level. Among these algorithms Decision tree algorithm provides 84% accuracy.

Keywords: Arrhythmia Diseases, Prediction, Machine Learning, KNN, Naïve Bayes, Decision tree, Accuracy

I INTRODUCTION

Changing the food habit is the foremost reason of some diseases like Heart attack, Blood pressure and cholesterol; here heart disease is the most dangerous disease. Age factor is also a one of the major argument here. So it is very significant to predict the heart related diseases in earlier. Computing concept are applied in various fields, in healthcare field machine learning techniques are used to guess the disease in the initial stages. These algorithms are used to detect the heart diseases automatically because the detection of the heart related diseases in the earlier manner decreases the death rate. In this research paper the main three machine learning concepts used to predict the heart diseases.

The main theme of the proposed system is to predict arrhythmia syndrome accurately with less number of clinical tests and to prevent and decrease cost for surgical treatment.

The other part of the paper ordered as: second section discuss the various techniques to predict the heart diseases and third section deals with the three machine learning concepts. The fourth section of this article describes the proposed architecture. Fifth section of the articles demonstrates the performance of classifiers in terms of accuracy level. Six section concludes the article.

II LITERATURE REVIEW

Data mining concepts are used to retrieve the information from the collection of data. This retrieved data can be used to take a better decision in business. In medical domain data mining concepts are used to discover important patterns used for medical diagnosis. Anjan Nikhil Repaka et al., uses Navies Bayesian algorithm to attain a better heart disease diagnosis result. The major parameters for diagnosing heart diseases are blood pressure, diabetes, cholesterol etc. Here 80% data set can be used like a training data set and 20% of information can be used as a test data. This proposed approach consists of various stages : data capturing, user login, classify the data using Navies Bayesian method and secure the data using AES method. The output of this proposed system shows that Navies Bayesian method is used to assist in detecting risk factors in heart disease [1].

Pahulpreet Singh Kohli et al., L., say that machine learning concepts are used to improve detection level of automatic diagnosis system. Here those authors applied various classifiers on different data sets available in UCI repository for disease forecasting. e p-value test can be used in feature section phase[2].

Abderrahmane Ed-daoudy et al., proposes a novel disease detection system using apache Spark software. It consists of three major parts: streaming process, storage the data and visualize the data. Spark MLlib is used for data streaming process and to apply classification. Apache Cassandra is used to store the created data [3]. Mr.Santhana Krishnan.J developed a new system to diagnosis the heart diseases. The output of the proposed system shows the possibilities of heart related diseases. Here the system can be implemented by using python programming with decision tree and Naive Bayes concept[4].

Seyedamin Pouriyeh et al., created a system to predict the heart related disease and evaluate the performance level of various classification algorithms such as KNN, SVM, Naive Bayes and decision tree. Bagging and staking can apply to predict the classifiers. The final output shows that SVM classifier provides better result than other classifiers [5].

Rahma Atallah et al., presented a major ensemble models to predict the presence of heart problem. The mail goal of this proposed system is to create the confidence level among the patients and increase the diagnosis accuracy level. Finally this ensemble model with had voting created better performance of 90% accuracy [6].

III Machine Learning Algorithms

KNN Algorithm

It is a non parametric function used for the classification method and regression concept. This algorithm works by finding the distance between a query and all the examples in the data, choosing the particular number examples closes to the query these algorithm votes for most frequent label or average the labels.

In this research paper euclidean distance property is used. Query Coordinate (a,b) and training samples is (c,d) then Square Euclidean distance is expressed as

$$X^2 = (c-a)^2 + (d-b)^2$$

KNN working can be explained on the basis of the following steps:

i. Select the number k of the neighbor.

ii. Compute the Euclidean distance value of K number of neighbor.

iii. Obtain the k nearest neighbor as per the computed Euclidean distance value

iv. Among from these k neighbors, calculate the number of the data point in each group.

v. allocate the new data points to that group for which the number of the neighbor is utmost.

Navie Bayes

In machine learning technique Naïve Bayes classifier is the relations of simple probabilistic classifiers based on using bayes theorem with strong autonomy assumptions between the features values. Where, Naïve Bayes classifier assumes that the presence of the concern feature in a class is unrelated to the presence of any extra feature data value.

$$P(A|B) = \frac{P(B|A)\ P(A)}{P(B)}$$

Decisions Tree

Decisions tree classifier is one of the predictive model approach is used in statistics, data mining technique and machine learning concept. This algorithm develops a decision tree to go from observation value about a starting entry to till target data. This algorithm comes under the family of supervised learning techniques. The main goal of using a decision tree algorithm is to generate a training model which can use to detect the class or value of the target variable by learning simple decisions rules referred from prior data. D is represented a sample data and the tree is developed depends on entropy value. Using divide and conquer approach the decision tree can be constructed easily. Remove the tree node by using pruning method.

$$Entropy = -\sum_{j=1}^{m} p_{ij} \log_2 p_{ij}$$

ALGORITHM

a) By using the training set to compute the entropy of current state variable H(S)
b) Each parameter, compute the information gain (IG) with respectively to the attribute ‘x’ represented by H(S,x).
c) Compute gain for each attribute ‘x’ by using Gain = Entropy – Information gain.
d) Choose the parameter which has greatest value of IG(S,x).
e) Take away the parameter that offers maximum IG from the group of attribute.
f) Do again until run out of all parameters, or the decision tree consists all leaf nodes.

IV Proposed System

In traditional heart related problem prediction requires lot of clinical tests. Precise and on time judgment of heart related disease is significant for heart failure avoidance and treatment. The healthcare organizations collect enormous amount of data that contains some unseen data, which is useful for building effective solution. To develop better results and building effective solutions on information, some machine learning concepts are used. Arrhythmia Disease Prediction System (ADPS) is developed with the help of K-Nearest Neighbor algorithm, Navies Bayes and Decision tree and for detecting the risk level value of heart related disease. The ADPS detects the probability of patients getting heart related disease. The following figure 1 describes the framework of proposed Arrhythmia disease prediction system.

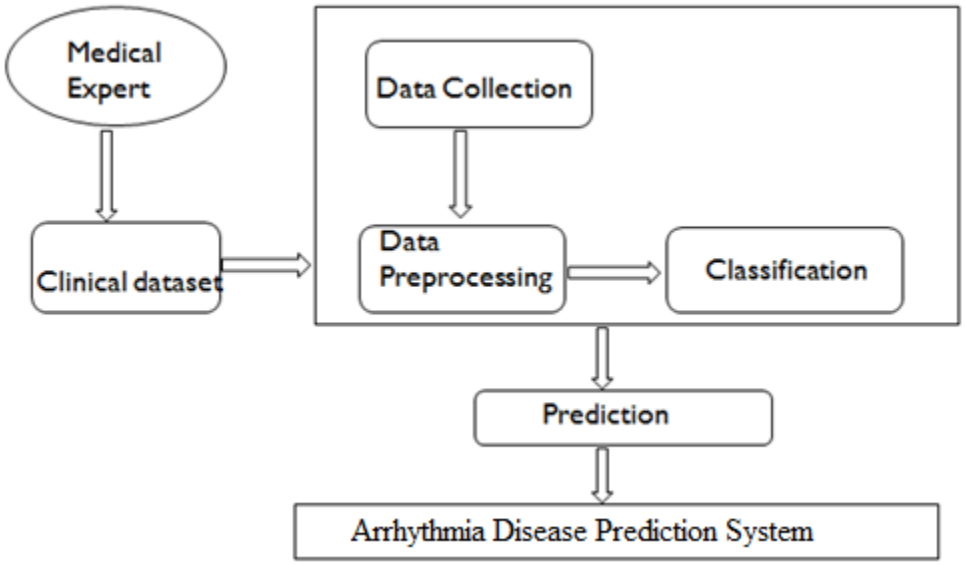

Figure 1 Developed ADPS Framework

The entire predictive model is composed of three stages namely collecting data, Preprocessing the data and Classification of data.

Data gathering:

First the information is gathering from the medical expert it refers to clinical dataset. Then the information is obtained from the collected dataset.

Data preprocessing:

In this preprocessing method the unwanted data is removed for the collected database, also if there are any missing values it can be replaced.

Classification:

In the preprocessed data all the three algorithms namely KNN, Decision tree and Naïve Bayes are used and the accuracy level is obtained.

After applying all the three algorithms, based on the accuracy level arrhythmia disease prediction model is obtained.

V RESULT AND DISCUSSION

Over the last decade almost one person dies of heart disease for a minute in USA. The earlier prediction of heart disease helps to safeguard the person for danger. In this proposed system we have applied the concepts of machine Learning Algorithms to forecast the heart diseases in earlier manner. In this research fourteen parameters are used to predict the arrhythmia disease. The important parameters are age, Cholesterol level, blood pressure level and sex. Based on the performance of the Algorithms accuracy level is measured.

The following table 1 shows the accuracy level of classification models used in this proposed system.

Sl. No	Classifier Name	Accuracy Level
1	Na¨ıve Bayes	78 %
2	KNN	80%
3	Decision Tree	84%

The following Figure 2 demonstrates the performance of proposed system. This system can be implemented using python programming.

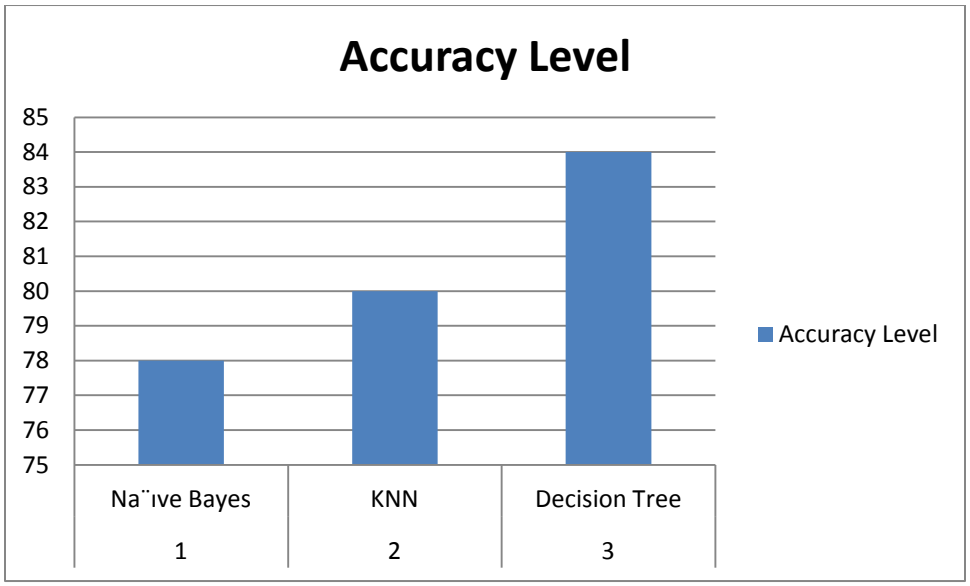

Figure 2 Accuracy Rates of Classification Models

VI CONCLUSION

The well known disease which affects human in middle or old age and leads to critical situation is heart disease. Classification is one of the best methods used to predict the diseases. Here the algorithms used to predict the heart disease are KNN algorithm, Naïve Bayes and Decision tree. The main advantage of decision tree classifier is the accuracy level and the characteristics of this algorithm is advanced than other methods, it is easy to understand, the rules are easily generated. Based on the accuracy level the best algorithm is decision tree for handling medical data. This system is implemented by using python programming and tested with real time data set. This data set consists of 304 patients details with 14 parameters. In future the disease can be predicted with more accuracy level by advance machine learning concept.

REFERENCES

[1] Anjan Nikhil Repaka, Sai Deepak Ravikanti & Ramya G Franklin (2019), " Design And Implementing Heart Disease Prediction Using Naives Bayesian", 3rd International Conference on Trends in Electronics and Informatics (ICOEI), pp. 292-297.

[2] Pahulpreet Singh Kohli & Shriya Arora (2018), " Application of Machine Learning in Disease Prediction, 4th International Conference on Computing Communication and Automation, pp. 1-4.

[3] Abderrahmane Ed-daoudy & Khalil Maalmi . (2019), " Real-time machine learning for early detection of heart disease using big data approach", International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS).

[4] Mr.Santhana Krishnan.J Dr.Geetha.S (2019), "Prediction of Heart Disease Using Machine Learning Algorithms", 1st International Conference on Innovations in Information and Communication Technology (ICIICT).

[5] Seyedamin Pouriyeh , Sara Vahid, Giovanna Sannino, Giuseppe De Pietro , Hamid Arabnia & , Juan Gutierrez (2017, " A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease", IEEE Symposium on Computers and Communications (ISCC).

[6] Rahma Atallah & Amjed Al-Mousa (2019), " Heart Disease Detection Using Machine Learning Majority Voting Ensemble Method", 2nd International Conference on New Trends in Computing Sciences (ICTCS).

[7] SENTHILKUMAR MOHAN , CHANDRASEGAR THIRUMALAI1 & GAUTAM SRIVASTAVA ," Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques ", Vol. 7, pp.81542 – 81554.

[8] S. Radhimeenakshi(2018), "Classification and prediction of heart disease risk using data mining techniques of Support Vector Machine and Artificial Neural Network", 3rd International Conference on Computing for Sustainable Global Development (INDIACom), IEEE.

[9] B. Bavani, S. Nirmala Sugirtha Rajini, M.S. Josephine & V.Prasana Kumari(2019), " Heart Disease Prediction System based on Decision Tree Classifier", Jour of Adv Research in Dynamical & Control Systems, Vol. 11, No. 10.