© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 12 Issue: 04 | October 2021
Paper is available at http://www.jrps.in | Email : info@jrps.in
Peer Reviewed and Refereed Journal

Heart Disease & Diabetes Prediction using Machine Learning

Kamal S.Chandwani[1]
Bhabha Engineering & Research Institute,Bhopal,M.P
chandwani1@rediffmail.com

Monika Raghuwanshi[2]
Bhabha Engineering & Research Institute,Bhopal,M.P.
monipriya21@gmail.com

***Abstract*—Over the last decade heart disease is the main reason for death in the world. Almost one person dies of Heart disease about every minute in India alone. In order to lower the number of deaths from heart diseases, there has to be a fast and efficient detection technique. Decision Tree is one of the effective data mining methods till this date. The algorithm used in this project is namely are Decision Tree, Naïve Byes, Support vector machine(SVM), k-nearest neighbours algorithm (KNN), Logistic regression, Random Forests. Heart disease defines several healthcare conditions that are vast in nature which is related to the heart and has many basic causes that affect the entire body.The data set employed in most of the concerned literature is Pima Indian Diabetic Data Set. Early diabetes detection is significant as it helps to reduce the fatal effects of the diabetes. Various machine learning techniques likeartificialneuralnetwork,principal component, decision trees, genetic algorithms, Fuzzy logic etc. have been discussed and compared. This paper first introduces the basic notions of diabetes and then describes thevarious techniques used to detect it. Anextensive literature survey is then presented with relevant conclusion and future scopes with analysis have beendiscussed.**

Keywords: Machine Learning, AI algorithms, Heart Attack, Cardiovascular Disease, Cleveland Dataset, Smart monitoring system, Fuzzy Logic, Fuzzy C-Means, SVM, GA, PCA, ANN.

I.INTRODUCTION

Heart disease is the kind of disease which can cause the death. Each year too many peoples are dying due to heart disease. Heart disease can be occurred due to the weakening of heart muscle. Also, the heart failure can be described as the failure of heart to pump the blood. Heart disease is also called as coronary artery disease (CAD). CAD can be occurred due to insufficient blood supply to arteries.Heart disease can be detected using the symptoms like: high blood pressure, chest pain, hypertension, cardiac arrest, etc. There are many types of heart diseases with different types of symptoms. Like: 1) heart disease in blood vessels: chest pain, shortness of breath, pain in neck throat, 2)heart disease caused by abnormal heartbeats:slow heartbeat, discomfort, chest pain., etc. Most common symptoms are chest pain, shortness of breath, discomfort, chest pain, etc. Most common symptoms are chest pain, shortness of breath, fainting. Causes of heart disease are defects you're born with, high blood pressure, diabetes, smoking, drugs, alcohol. Sometimes in heart disease the infection also which affects the inner membrane which is identified by symptoms like fatigue, dry cough, skin rashes. Causes of heart infection are bacteria, viruses, parasites. Types of heart disease: Cardiac arrest, Hypertension, Coronary artery disease, Heart failure, Heart infection, Congenital heart disease, Slow heartbeat, Stroke type heart disease, angina pectoris. Now a days therearetoo many automated techniques to detect the heart disease like data mining, machine learning, deep learning, etc.

ML plays a very important role to detect the hidden discrete patterns and thereby analyze the given data. After analysis of data ML techniques help in heart disease prediction and early diagnosis.In this we train the datasets using the machine learning repositories. There are some risk factors on the basis of that the heart disease is predicted. Risk factors are: Age, Sex, Blood pressure, Cholesterol level, Family history of coronary illness, Diabetes, Smoking, Alcohol, Being overweight, Heart rate, ChestPain. Machine Learning is a branch of AI research [2] and has become a very popular aspect of data science. The Machine Learning algorithms are designed to perform a large number of tasks such as prediction, classification, decision making etc.

One of the most important organs of the human body is a heart. one of the most common cardiac diseases in India is the heart attack. The heart pumps blood through the circulatory system of the body. In all body part the blood, oxygen is circulated by the circulatory system of the body and if the heart does not work properly then the whole human blood systemwill be collapsed.So if the heart does not function properly then it will lead to a serious health condition, it could even lead todeath.

1.1 Types of the Heart Disease: CardioVascular disease (CVD) or also known as heart disease include blood and heart of the human body. myocardial infarction (as a heart attack) is also apart of the CVD. Another type of Heart Disease is called Coronary Heart Disease(CHD), in this type of disease, a substance called Plaque develop in the coronary arteries. The development of plaque can block the vessel completely through the course of time. Diabetes is one of the diseases that are spreading like epidemics in the entire world. It is seen that every generation ranging from children, adolescents, young people and old age are suffering from it. Pro-long effect can cause worse effects in terms of failure of organs like liver, kidneys, heart, stomach and can lead to death. It is frequently associated with the disorders-Retinopathy andNeuropathy.

The symptoms of the Heart Attack :

1. Chest Pain: The most common sign of a heart attack is chest pain. It

mainly happens cause of the blockage of the coronary vessel of the body due to theplaque.

2. Arms pain: The pain normally starts in thechest and move towards the arm mainly leftarm.

3. Low in oxygen: Because of the plaque the level of oxygen drops in the body and causes the dizziness and loss ofbalance.

4. Tiredness: this cause for fatigues means simple chores become harder to do.

5. Excessive Sweating: Another common symptom issweating.

6. Diabetics: In this, the patients have a heart rate of ~ 100 bpm and also occasionally having a heart rate of130bpm.

7. Bradycardia: In this, the patient will have a slower heartbeat of 60bpm.

Cerebrovascular Disease: The patient will have a high heart rate than normal usually of200bpm and higher than this can cause a Heartattack.

8. Hypertension: In this the patient's heart rate normally ranging from 100-200bpm.

Diabetes is mainly of two types-type 1 and type 2.

Type -1 Diabetes :- It is the situation in which liver does not produce insulin at all. Insulin is an hormone that is required to absorb glucose from the blood to utilize this glucose for body building. However, absence of insulin in the body will increase blood sugar and it will lead to it. It is commonly found in children and adolescents. It mainly occurs because of the genetic disorders. It is often known as juvenile disorder. Its common symptoms are frequent urination. Weight loss, increases thirst, blurs vision, nerves problems. This can be treated by insulintherapy.

Type -2 Diabetes :-It is long term metabolic disorder generally occurs in the adults over age of 40 years. It is evident by high blood sugar, insulin resistance and high insulin. The major cause is obesity and lack of exercise. This bad lifestyle can cause glucose to get store in the blood and develop diabetes.90% of people affected by type-2 diabetes only. To treat insulin resistance metformin is given to ensure this can betreated.

Diabetic Neuropathy :- These are the nerve disorders developed in diabetic patients with the passage of time. They often occur in foot and hands. The common symptoms are pain, numbness, tingling, loss of feeling in hand, foot, arms etc.

Diabetic Retinopathy :- It is the diabetic disorder that leads to permanent eye blindness. Initially there is no significant symptom, gradually symptoms are seen. In the second stage, bloodvessels are developed at the back of the eyes that could lead to bleeding on bursting as they are quite agile.

II. Proposed Work:

In this paper, comparison of various machine learning methods is done for predicting the 10 year risk of coronary heart disease of the patients from their medical data. The following is the flowchart for proposed methodology:

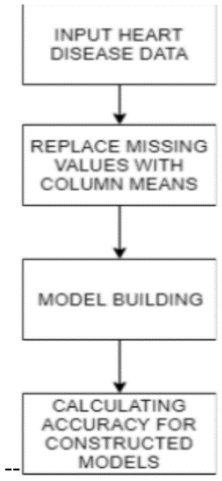


FIGURE 1: PROPOSEDWORK

The heart disease data set is taken as input. It isthen pre-processed by replacing non- availablevalues with column means. Three different methods were used in this paper.The output is the accuracy metrics of the machine learning models. The model can then be used in prediction.

III.. Various Methodologies**:**

Naive Bayes Classifier: It is a classifier technique based on "Bayes theorem", it assumes a particular class which has a particular feature is unrelated to other features in class. We used it in the model because it is easy tomake and useful for large dataset. It provides data structures such as Network Structure, Conditional probability distribution and others. Figure 3 shows the naive Bayes code.

```
#Gausssian Naive Bayes
model = train_model(X_train, y_train, X_test, y_test, GaussianNB)

Train accuracy: 85.38%
Test accuracy: 86.81%
```

Figure 2 .Naive Bayes Train and Test accuracy

Support Vector Machines(SVM): A Support Vector Mechanism (SVM) is a discriminative classifier formally defined by a separating hyperplane. Simply put the algorithm outputs an optimal hyper plane which categorises new examples. In 2-D space, this hyper plane is a line dividing a plane into two parts wherein each class lay on either side. The points which positions are in the separating Hyperplane is called Support Vectors. The distance between the Canonical and Separating hyperplane is called Margin. We have implemented a variant of SVM which is sequential minimal optimization. The minimal sequential optimization breaks the problems in subproblems and then solves itanalytically.

Logistic Regression: Logistic Regression is a method which analyses a dataset which has a one or more independent variable and gives an outcome. The goal of the Logistic Regression is to predict the best relationship between the dependent and independent variables. Figure 3 showsthe Logistic Regression of the model and the accuracy of the test and train model.

```
# Logistic Regression
model = train_model(X_train, y_train, X_test, y_test, LogisticRegression)
```

```
Train accuracy: 85.85%
Test accuracy: 85.71%
```

Figure 3. Logistic Regression

Decision Tree: Decision tree is used for making a tree like structures for regression or classification models. A decision tree creates a smaller and smaller subset of a problem while an associated decision tree is developed incrementally. Two or more branches and leaf can seem in a decision tree which represents classification. Both categorical and numerical value can be handled by a decision tree.Thealgorithm Decision tree can learn to predict the value of a target variable by learning simple decision rules taken from the dataset. From the result of our decision tree, we can easily understand how much importance a particular feature has. In figure 5 we can see the feature 'Thal' is turned out to be a very important feature of our model.

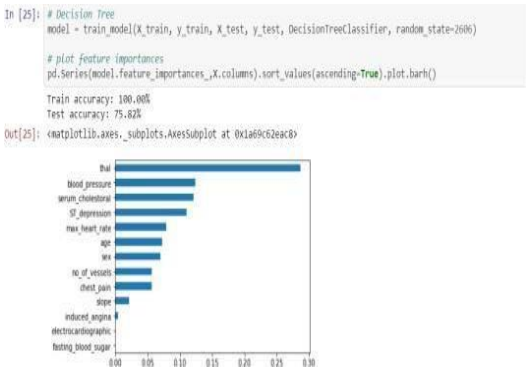


Figure 4. Domain tree

Here the decision tree learns the train set model perfectly and overfitting the data. That's why it will give a poor prediction. Other values of 'max_depth' parameter need.to be tried out, it is shown in Figure

Figure 5 'max_depth' parameter With

Random forest: Random Forest algorithm does not overfit the set like 'Decision Tree'. Random Decision Tree first considers many decision trees before giving an output. Random forest algorithm uses a voting system for classification where it decides the class. It works well with the biggerdataset.'max_depth' six the score went to almost 80% of the decision tree.

K-Nearest Neighbour (KNN):KNN is a supervised classification algorithm (it takes a bunch of labeled points and uses them how to label another point).To label a new point it looks at the new point nearest to it and votes for it and whichever label is the most voted that label is givento the new point. Below in figure 5 we can see KNN model

```
# KNN
model = train_model(X_train, y_train, X_test, y_test, KNeighborsClassifier)
```

```
Train accuracy: 88.21%
Test accuracy: 86.81%
```

Figure 5 .KNN train and test accuracy

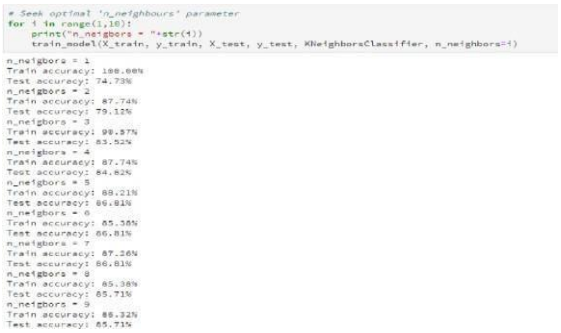


Figure 6 'n_Parameter'

Despite its simplicity, the result is very good so we put different values forn.

Figure 6.RandomForest

Fuzzy C-means

It is an extension of K-means clustering algorithm that means it aims at forming the clusters, then finding out thecentroidsof the clusters, the incoming data set is assigned to that cluster that has minimum distance from it's centroid.However,it may happen that sometimes very less margin is there so that new data set can be fall for more than one cluster. This was avoided by fuzzy C-means clustering algorithm as it employs fuzzy partition that accounts for the membership function.Hence,results produce are

more accurate.

Principal Component Analysis

PCA is a statistical model that is used to classify data set in such a way that the maximum co- relation can be found in the data set. It aims at construction to orthogonal plane so that data can be classified along with this plane, another plane is perpendicular on it, that is known for second co-relation among data set. It helps in feature extraction and makes use of Eigen values and Eigen vectors to calculate the principal component.

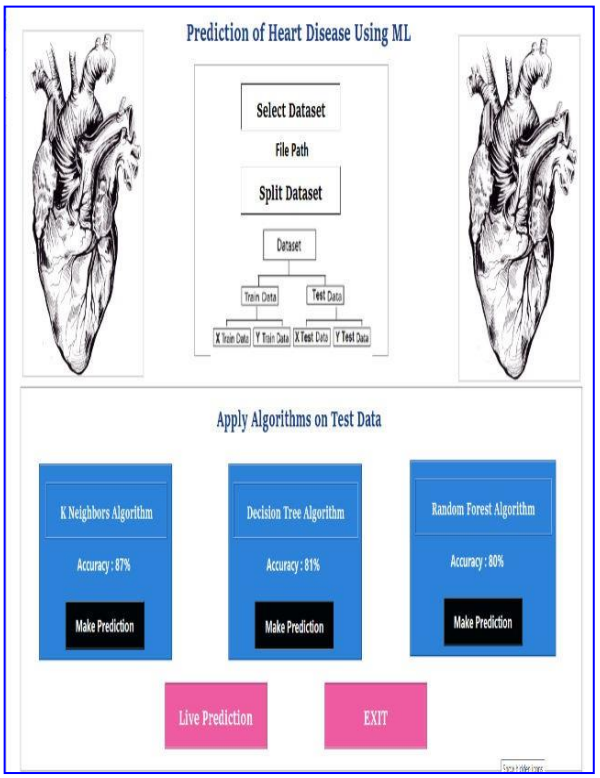
IV. Modules Used:

Login Module

In This module User able to login into the system to access the other services.Login Module is a portal module that allows users to log in. You can add this module on any module tab to allow users to log in to thesystem.

Algorithm Comparison Module

In this module we are going to compare three different (Random Forest Algorithm, Decision tree and k- nearest neighbour) and plot into chart. In this module we are going to check prediction on bulk amount of different user data and check the Accuracy ofalgorithm.



Live Heart Disease Prediction Module

- In this module user can able to predict whether he/she has heart disease or not by entering all attributes Value into the interface and get the prediction from thebest algorithm that was KNN where we achieves the 87 %accuracy.

The accuracy of the algorithms is calculated. Theaccuracy results are tabulated as follows:

Method	Method Accuracy
Decision tree	81.00 %
k- nearest neighbour	87.00 %
Random Forest Algorithm	80.77 %

The accura

Accuracy of K-nearest neighbor algorithm isgood when compared to other algorithms and it is best algorithm that fit in this kind of problem.

VI Conclusion & Future Scope:

In this paper, we have presented a system which is suitable for real-time heart diseases prediction and can be used by the users who have coronary disease. Different from many other systems it is able to both monitor and prediction. The diagnosis system of the system is able to predict the heart disease by using ML algorithms and the prediction results are based on the heart disease dataset instance. On the other hand, the system is very inexpensive, we used amped pulse sensor and send the data to mobile via Arduino suite microcontroller. For checking the variances and raise the alarm if the user's heart rate rise than the normal rate of the heart. To prove the effectiveness of the system we have carried out experiments for both monitoring and diagnosis system .we ran experiments with some popular algorithms like KNN, Decision Tree, Random Forest, Naive Bayes, SVM, Logistic Regression. The experiment was carried out with the holdout test and the accuracy of the proposed system was 89% achieved with the Randomforest.

REFERENCES

[1] Senthilkumar Mohan, ChandrasegarThirumalai, GautamSrivastava —Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques‖, Digital Object Identifier 10.1109/ACCESS.2019.2923707, IEEE Access, VOLUME 7,2019 S.P. Bingulac, —On the Compatibility of Adaptive Controllers,‖ Proc. Fourth Ann. Allerton Conf. Circuits and Systems Theory, pp. 8-16, 1994. (Conference proceedings)

[2] SonamNikhar, A.M. Karandikar" Prediction of Heart Disease Using Machine Learning Algorithms" International Journal of Advanced Engineering, Management and Science (IJAEMS) Infogain Publication,[Vol-2, Issue-6, June- 2016].I.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[3] AditiGavhane, GouthamiKokkula, IshaPandya, Prof. Kailas Devadkar (PhD)," Prediction of Heart Disease Using Machine Learning", Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018).IEEE Conference Record # 42487; IEEE Xplore ISBN:978-1- 5386-0965-1

[4] Abhay Kishore1, Ajay Kumar2, Karan Singh3, Maninder Punia4, Yogita Hambir5," Heart Attack Prediction Using Deep Learning", International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 04 |Apr-2018.

[5] A.Lakshmanarao, Y.Swathi, P.SriSaiSundareswar," Machine Learning Techniques For Heart Disease Prediction", International Journal Of Scientific & Technology Research Volume 8, Issue 11, November2019.

[6] Mr.SanthanaKrishnan.J, Dr.Geetha.S," Prediction of Heart Disease Using Machine Learning Algorithms",2019 1st International Conference on Innovations in Information and Communication Technology(ICIICT),doi:10.1109/ICIICT1.2019.8741465.

[7] AvinashGolande, Pavan Kumar T," Heart Disease Prediction Using Effective Machine Learning Techniques", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1S4, June2019.

[8] V.V.Ramalingam,AyantanDandapath,MKarthikRaja,"Heartdisease prediction using machine learning techniques: a survey", International Journal of Engineering & Technology, 7 (2.8) (2018)684-687.

[9] . Manikantan and S. Latha, "Predicting the analysis of heart disease symptoms using medicinal data mining methods", International Journal of Advanced Computer Theory and Engineering, vol. 2, pp.46-51, 2013.

[10] M. S. Amin, Y. K. Chiam, K. D. Varathan,''Identification of significant features and data mining techniques in predicting heart disease,''

Telematics Inform., vol. 36, pp. 82–93,Mar.2019.

[11] S. M. S. Shah, S. Batool, I. Khan, M. U. Ashraf, S. H. Abbas,and S. A. Hussain,‘‘Feature extraction through parallel probabilistic principal component analysis for heart disease diagnosis,’’ Phys. A, Stat.Mech.Appl.,vol. 482, pp. 796–807,2017.doi:10.1016/j.physa.2017.04.113.

[12] Stephen F. Weng, Jenna Reps, Joe Kai1, Jonathan M. Garibaldi, NadeemQureshi,—Can machine-learning improvecardiovascular risk prediction using routine clinical data?l, PLOS ONE | https://doi.org/10.1371/journal.pone. 0174944 April 4,2017.

[13] N. Al-milli, __Backpropagation neural network for prediction of heart disease, ‘‘J. Theor. Appl.Inf. Technol., vol. 56, no. 1, pp.131–135, 2013.

[14] A. S. Abdullah and R. R. Rajalaxmi, ‘‘A data mining model for predicting the coronary heart disease using random forest classifier,’’ in Proc. Int. Conf. Recent Trends Comput. Methods, Commun. Controls, Apr. 2012, pp.22–25.