

Early Detection of Heart Syndrome Using Machine Learning Technique

Noor Basha
Department of CSE
Vemana Institute of Technology,
Bengaluru, India
md.noor204@gmail.com

Gopal Krishna C
Department of CSE
Adi Chunchanagiri Institute of Technology,
Chikmagaluru, India
nischal.gopal@gmail.com

Ashok Kumar P S
Department of CSE
Don Bosco Institute of Technology,
Bengaluru, India
ashokydit@gmail.com

Venkatesh P
Department of TCE
Don Bosco Institute of Technology,
Bengaluru, India
Venkateshp@gmail.com

Abstract – Analysis and Prediction of diseases are two most demanding factors to be faced critically by the doctors and data scientist, where data analytics be very delightful issue, so in this regard, many health industries will working on variety of human syndromes, where they generate huge data. Heart disease, cancer, tumour and Alzheimer’s disease are one of the chronic human diseases, where data scientist and doctors are doing rapid and efficient analysis on these diseases using many machine learning techniques to study and predict these diseases to save and reduce human deaths.

Importance of this article is to predict and analyze the heart related syndrome in patients, based on one of the main feature, like age, where data scientists can do predictive research on big data to early analysis on heart syndrome to save the life of the patients. In this case study many features are well thought-out to do AN analysis and predict of heart diseases in patients, here author checked with prediction of data using many machine learning algorithm are used to verify the performance of syndrome

Keywords–Heart Disease, KNN, Decision Tree, Random Forest, SVM, Naive Bayes

I. INTRODUCTION

Due to busy schedule as well as routine assignments peoples are facing severe stress and anxiety. More over some other peoples are addicted with chronic habitual behaviour, like consumption of Cigars and Gutuka, those peoples are suffering from chronic diseases like, heart diseases, cancer, Liver problems, Kidney failures etc. To cure such persons with chronic disease is a big hurdle to well know doctors, is a current world issue. Regarding this new challenge, IT professionals are provided hand to hand support to predict such disease early and cure as well as recover the patients from the chronic disease.

A. Heart Disease – Case study

In the present scenario each humans are so exceptional in his individual features and manners, but even though every humans may have different pulse rate as well as blood pressure ratings. Based on the history and generic evaluation of medical practitioners and researchers believed that, a healthy humans pulse rate is varied in between of 60 to 100 bpm and BP is varied in between of 120/80 to 140/90 (mm Hg), and these readings are proved by medical practitioners.

Heart syndrome is one the vital abrupt death or accidental death of humans in this world, this is might be

because of poor dieting as well as physical exercise and other activities like, consumption of alcoholic products, smoking etc. In this article, author is tried to predict and analysis the heart syndrome with respect to many features like age, gender, blood pressure, heart rate, diabetes etc, but however actual prediction of heart syndrome is totally a critical task to the medical practitioners and analyst.

In present market, health industries has many machine learning tools and techniques are used to predict various chronic diseases, but still researchers find some sort of flaws, so they expect some more effective and efficient predictive algorithms to find chronic diseases of humans in early stage itself, so that we can save the life of the patients.

II. REVIEW OF LITERATURE

Coronary heart disease narrows down the coronary arteries. Basically coronary arteries will supply both oxygen as well as blood to heart, if the heart functioning is not proper, and then it causes to malfunctioning of heart leads to ill or death to a person [13].

J Thomas, R Theresa Princy [2] made use of many classification algorithms to predict the severe heart syndromes based on risk rate, where the author specifically used data mining approach.

Sana Bharti, Shailendra Narayan Singh [3] made use of artificial neural network and Genetic algorithm to predict health diseases. In this reference, author collaborate the data mining approach with association rules and classification techniques. In this regard the model developed by the author is so efficient on predicting the heart syndrome.

Himanshu Sharma, M A Rizvi [6] made use of various classification algorithms to create effective data analysis model on prediction of severe heart syndromes. Usually dataset may contain noise features, it abruptly corrupt the valid data, so they tried to reduce the noise by cleaning and pre-processing the dataset and also tried to reduce the dimensionality of the dataset. They found that good accuracy can be achieved with neural networks.

Animesh Hazra et.al, [7] discussed in detail the cardiovascular disease and different symptoms of heart attack. The different types of classification and clustering algorithms and tools were used.

V.Krishnaiah, G.Narsimha, N.Subhash Chandra [8] presented an analysis using data mining. The analysis showed that using different techniques and taking different number of attributes gives different accuracies for predicting heart diseases.

Ramandeep Kaur, Er.Prabhsharn Kaur [9] have showed that the heart disease data contains unnecessary, duplicate information. This has to be pre processed. Also, they say that feature selection has to be done on the dataset for achieving better results.

III. METHODOLOGY

Figure 1 signifies the proposed architecture of the learning model, where many machine learning algorithms are used to check and predict the severe heart syndrome in the critical heart patients. In this model Heart Disease data from the kaggle web site is to be considered as a input data and process the data.

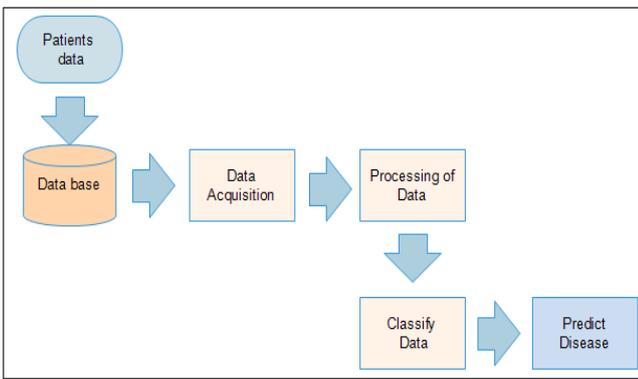


Fig 1: Proposed Architecture of the system

If person is suffering from acute diabetic from very long period, then sure in future such person may expected very high probability of chronic heart syndrome, i.e. just visualize if person is acute diabetic and obsessed with smoking and alcoholic contents will sure it develop acute heart syndrome threat.

A. Data sources

The dataset used in this article is fetched from Kaggle web, basically Kaggle supports various dataset with open source and accessible with various data format more over data is supported with all platform.

TABLE 1. HEART_DISEASE.DATA SET

| SI No. | age | sex | cp | trestbpsm | chol | fbs | Restecg | thalach | exang | oldpea kST | slope | ca | thal | target |
|--------|-----|-----|----|-----------|------|-----|---------|---------|-------|------------|-------|----|------|--------|
| 0 | 40 | 1 | 2 | 140 | 289 | 0 | 0 | 172 | 0 | 0 | -9 | -9 | -9 | 0 |
| 1 | 49 | 0 | 3 | 160 | 180 | 0 | 0 | 156 | 0 | 1 | 2 | -9 | -9 | 1 |
| 2 | 37 | 1 | 2 | 130 | 283 | 0 | 1 | 98 | 0 | 0 | -9 | -9 | -9 | 0 |
| 3 | 48 | 0 | 4 | 138 | 214 | 0 | 0 | 108 | 1 | 1.5 | 2 | -9 | -9 | 3 |
| 4 | 54 | 1 | 3 | 150 | -9 | 0 | 0 | 122 | 0 | 0 | -9 | -9 | -9 | 0 |
| 5 | 39 | 39 | 1 | 3 | 120 | 339 | 0 | 0 | 170 | 0 | 0 | -9 | -9 | 0 |

The Data source heart_desiase.csv accomplished with various features like, age, sex, Cp, chol, FBS, Restecg, Thalach, Exang, Old Peak, Slope, Ca, Thal, where data set is used to build an efficient model to classify and predict the severity of heart diseases in critically suffered patients shown in table 1.

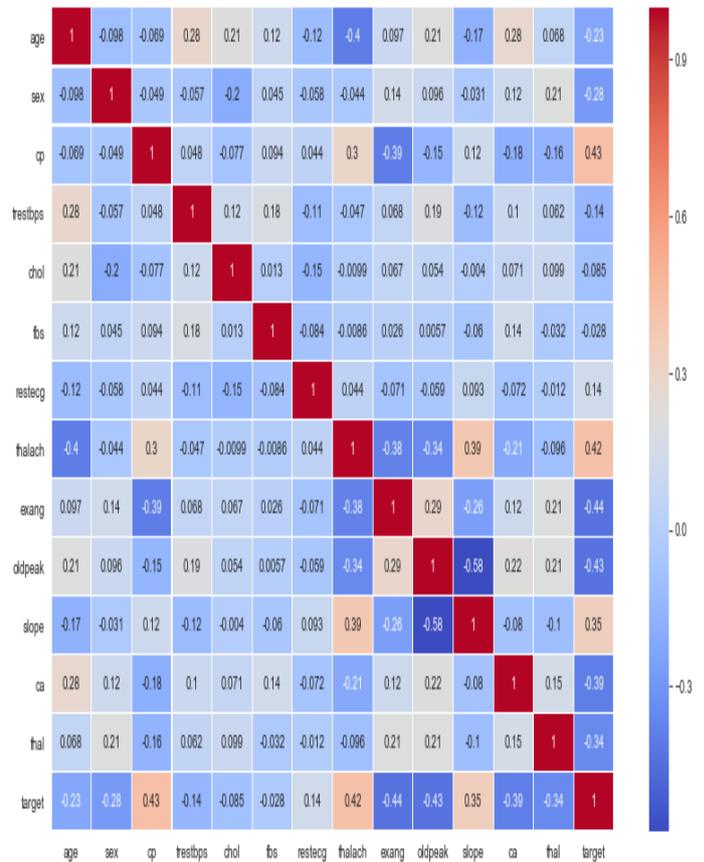


Fig 2: exploratory data analysis of Heart Disease data

As we aware that, high BP is one of the big warning sign to such persons, where heart is working too hard to pump the blood to whole body and it is obvious that it leads stress as well as strain to the heart it damages the blood veins as well as blood vessels.

Usually senior citizens and hereditary patients are easily infected by chronic heart syndrome, in this regard some of the healthy persons feature like, age, gender, poor or unhealthy diet and stress in work, etc.

Now a day at present world, working peoples or corporate employees are the potential victim community by the massive heart disease patients.

Figure 2, represents the exploratory data analysis of Heart Disease data set. An exploratory data analysis (EDA) is an approach to analyzing data sets to summarize their main characteristics with visual methods. Primarily EDA is for seeing what the data can tell us beyond the formal modelling or hypothesis testing task, where EDA explore the data and formulate the hypothesis.

Figure 3 represents the sample graph of heart syndrome on patients with various features, if a person is doing very high stressful work and adopted poor dieting personality, it damn sure that such person will get infected within a short period of years and he is one of the potential victim of severe coronary attack.

In this regard many researchers and creative medical practitioners are doing critical analysis on heart syndrome and other chronic disease symptoms to detect in early stages itself to save the patients health from these health hazards, so researchers are using many kinds of learning algorithms and techniques to achieve accuracy and efficiency on analysis and prediction of heart syndromes.

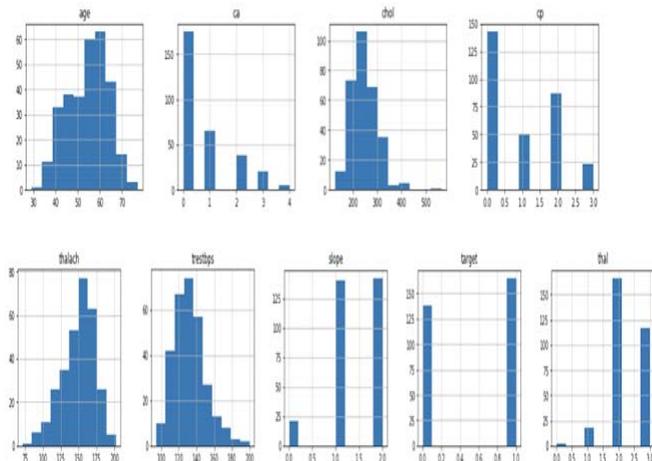


Fig 3: prediction of heart disease based on features

A. Machine Learning Algorithms

Machine learning is a kind of system learning procedure, where system will get an ability to operate automatically or itself, by proper training, so it improve system operation and experience efficiently without human intervention, i.e. the emphasis of machine learning is to develop an effective programs that to access data and its usage automatically.

Machine learning is to provide effective training on data set based on effective learning algorithm, where algorithms will create set of rules and responsibilities, but these rules are based on the inferences on data. While training the system, various dataset could be used with same learning algorithm to create multiple system models.

A. K Nearest Neighbors Classifier

KNN classifier identifies the classes for K nearest neighbours of a given data point to assigns a class, where variation of the neighbours within a class. For the calculation of test score, where the test scores are varied from 1 to 20 neighbours.

B. Decision Tree Classifier

In the decision tree classifier, it creates a tree based on which class values is going to assign on each data point, here for the effective data model creation, we acquired and considered maximum features to creating the efficient model. The features are ranging from 1 to 30 features.

C. Random Forest Model

Random forest model consist set of individual decision trees, that operate as collection, where each tree in the random forest split out into class prediction. The Random Forest method introduces more randomness and diversity by applying the bagging method to the feature space. That is, instead of searching greedily for the best predictors to create branches, it randomly samples elements of the

predictor space, thus adding more diversity and reducing the variance of the trees at the cost of equal or higher bias. This process is also known as “feature bagging” and it is this powerful method what leads to a more robust model

D. Naive Bayes

Naive Bayes is a statistical classifier, which assumes no dependency between attributes. It classifies conditional independence, where it assume an attribute value on a given class but it is independent on the values compare to other attributes values.

E. Support Vector Machine (SVM)

SVM is a supervised learning approach, which performs a similar task like C4.5 algorithm, except that it doesn't use Decision trees at all. Support vector machine attempts to decrease any chance of misclassification

IV. IMPLEMENTATION DETAILS

Above specified machine algorithms are used this dataset to complete the implementation, where KNN algorithm has very high accuracy compare to other algorithms, shown in below result analysis.

proposed machine learning model is tested and verified on heart disease data set, where it identifying the accuracy of above algorithms. Initially we referred confusion matrix, which describes performance of a classification model, where matrix contains information about actual and predicted data classifications performed by a classifier, shown in table 3.

As we aware that TP, TN, FP, FN are four major parameters used in confusion matrix, in which the actual and predicted values are defined below,

True positive (TP) = 39
 True negative (TN) = 38
 False Positive (FP) = 5
 False negative (FN) = 6

Where, the sensitivity and specificity can be defined as,

$$\text{Sensitivity} = 100 * TP / (TP+FN) = 100 * 39 / 45 = 86.66\%$$

$$\text{Specificity} = 100 * TN / (FP+TN) = 100 * 38 / 43 = 88.37\%$$

Table3: Confusion Matrix For Classification Of Heart Disease.

| Prediction | Actual | |
|------------------|---------------|------------------|
| | Heart disease | No heart disease |
| Heart disease | 39 | 5 |
| No Heart disease | 6 | 38 |

In order to show the efficiency of the proposed approach, other data records have been simulated and classified based on the model previously trained.

Figure 4, 5 & 6 represents the performance evaluation of the implementation of KNN, Decision Tree & random forest machine learning algorithms to be used for prediction of heart syndrome.

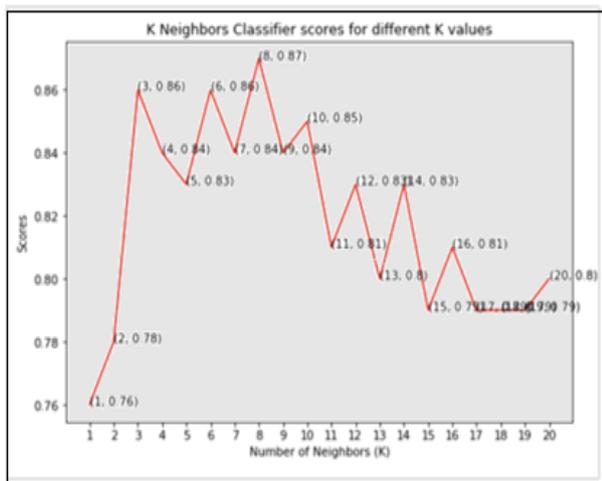


Fig 4: KNN Classifier scores for different features

Once we have the scores, we can then plot a line graph and see the effect of the number of features on the model scores.

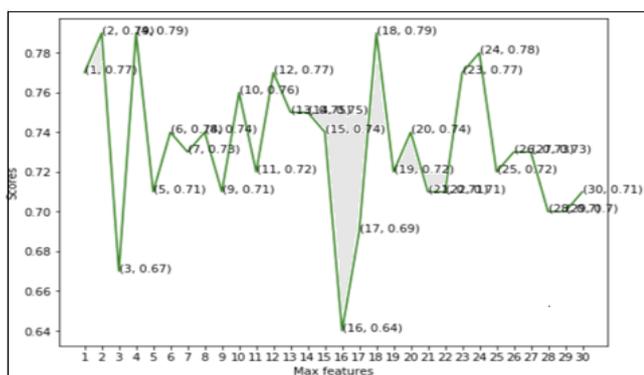


Fig 5: Decision tree Classifier scores for different features

In Random Forest classifier, it takes the concept of decision trees to the next level. It creates a forest of trees where each tree is formed by a random selection of features from the total features. Here, we can vary the number of trees that will be used to predict the class, where author calculate test scores over 10, 100, 200, 500 and 1000 features.

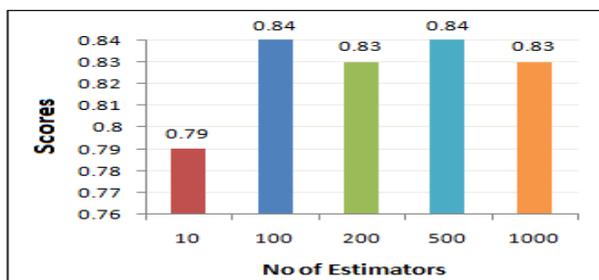


Fig 6: Random Forest Classifier Scores for different estimators

V. RESULT ANALYSIS

The overall result analyses of above algorithms are implemented and verified, where Figure 7 & 8 shows the Sensitivity, Specificity and Accuracy scores are delivered by all machine learning algorithms. Out of which KNN

algorithm has very good Accuracy score compared to other algorithms.

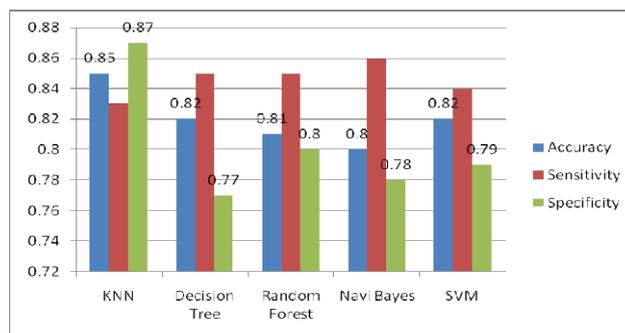


Fig 7: graph on Accuracy, Sensitivity, Specificity

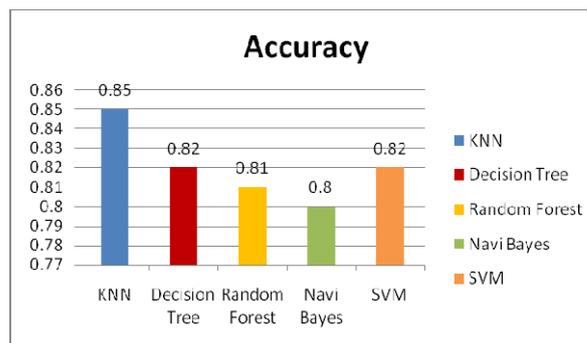


Fig 8: graph on Accuracy score

VI. CONCLUSION

Parental history or hereditary symptoms will leads to many chronic diseases to peoples, out of which heart disease is one among. If we identify the chronic diseases in early stage, it can be cured. So medical or hospital data set is collected from kaggle web to analyse and implement the data on different algorithm to check the accuracy score, sensitivity and specificity of the key attribute of the heart disuse patients.

In fact we analyse the proposed model for heart disease patients with various algorithm, in which many key attributes are verified, out of which KNN algorithm found to be that very effective and efficient performance on accuracy score on heart disease prediction.

With this inference of this customized model, machine learning algorithms will provide very valuable knowledge on analysis and prediction of many chronic diseases, so in this regard researchers are helpful to the needy persons, doctors and society.

REFERENCES

- [1]. Monika Gandhi, Shailendra Narayanan Singh Predictions in heart disease using techniques of data mining (2015)
- [2]. J Thomas, R Theresa Princy Human heart disease prediction system using data mining techniques (2016)
- [3]. Sana Bharti, Shailendra Narayan Singh, Amity university, Noida, India Analytical study of heart disease prediction comparing with different algorithms (May 2015)
- [4]. Purushottam, Kanak Saxena, Richa Sharma Efficient heart disease prediction system using Decision tree (2015)
- [5]. Sellappan Palaniyappan, Rafiah Awang Intelligent heart disease prediction using data mining techniques (August 2008)
- [6]. Himanshu Sharma, M A Rizvi Prediction of Heart Disease using Machine Learning Algorithms: A Survey (August 2017)

- [7]. Animesh Hazra, Subrata Kumar Mandal, Amit Gupta, Arkomita Mukherjee and Asmita Mukherjee Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review (2017)
- [8]. V.Krishnaiah, G.Narsimha, N.Subhash Chandra Heart Disease Prediction System using Data Mining Techniques and Intelligent Fuzzy Approach: A Review (February 2016)
- [9]. Ramandeep Kaur, 2Er. Prabhsharn Kaur A Review - Heart Disease Forecasting Pattern using Various Data Mining Techniques (June 2016)
- [10]. J.Vijayashree and N.Ch.SrimanNarayanaIyengar Heart Disease Prediction System Using Data Mining and Hybrid Intelligent Techniques: A Review (2016)
- [11]. Benjamin EJ et.al Heart Disease and Stroke Statistics 2018 At-a-Glance (2018)
- [12]. Abhay Kishore, Ajay Kumar, Karan Singh, Maninder Punia, Yogita Hambir Department of Computer Engineering, Army Institute of Technology, Pune,
- [13]. Manjunath K, Noor Basha, Ashok Kumar P S, Venkatesh P, Analysis and Forecasting of Heart Syndrome by Intelligent Retrieval Approach, ICTIDS-2019, Malaysia.
- [14]. S. Vijayarani et. al., "An Efficient Classification Tree Technique for Heart Disease Prediction",International Conference on Research Trends in Computer Technologies (ICRTCT - 2013) Proceedings published in International Journal of Computer Applications (IJCA) (0975 – 8887), 2013 (pp 6-9).
- [15] Harsh Vazirani et. al., " Use of Modular Neural Network for Heart Disease", Special Issue of IICCT Vol.1 Issue 2, 2010 for International Conference [ACCTA-2010], 3-5 August 2010 (pp 88-93).