International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 07 | July 2020 www.irjet.net p-ISSN: 2395-0072

Comparative Study of Cardiovascular Disease Detection Algorithms

Eqbal Ahmad[1], Arpan Kumar[2]

[1]*M. Tech Scholar, Bhabha Institute of Technology, Kanpur, UP (India)*
[2]*Assistant Professor, Dept. Of CSE, Himalayan Institute of Technology and Management, Lucknow, UP (India)*

Abstract - *The prognosis of cardiovascular disease is a sizable and tedious challenge in the field of medical treatment. The healthcare surrounding is commonly perceived as being 'Data Rich' yet 'Poor Knowledge'. There is a wealth of information to be had in the healthcare systems. In this field, there are lack of data analysis tools through which the data can be analysed and various disease can be detected. In the field of medical, a lot of information regarding disease can be extracted using various algorithms. Data mining could be a technique that's performed on large number of data for extraction of hidden patterns by using various strategies from statistical analysis, machine learning and database techniques. Further, the medical data processing is an especially important research field thanks to its importance within the development of assorted applications in flourishing healthcare domain. While summarizing the deaths occurring worldwide, the centre disease appears to be the leading cause. The identification of the chance of cardiovascular disease in a very person is complicated task for medical practitioners because it requires years of experience and intense medical tests to be conducted. In this work four data mining techniques are K-Nearest Neighbours, Support Vector Machine, Decision Tree and Random Forest algorithm is analysed and compared to find the best algorithm for cardiovascular disease detection.*

Key Words: **Cardiovascular Disease, Data Mining, K-Nearest Neighbours, Support Vector Machine, Decision Tree, Random Forest.**

1. INTRODUCTION

Heart diseases are referred to as cardiovascular diseases which occur thanks to unhealthy lifestyle, smoking, alcohol and high intake of fats which can cause hypertension, high pressure level, diabetics and strokes. The World Health Organization(WHO) [1] analysed that thirteen millions death worldwide was due to Heart diseases in 2017. In day to day life many factors that affect an individual's heart. Many problems are occurring at a high speed and new cardiovascular diseases are rapidly being identified. In today's world of stress, Heart, being a vital organ of the body which pumps blood through the body for the blood circulation is crucial and its health is to be conserved for a healthy living. The health of a person's heart is predicated on the experiences in an exceedingly person's life and is totally addicted to professional and private behaviours of someone. There can also be several genetic factors or family factors through which a sort of cardiovascular disease is passed down from generations.

In India casualties are moreover caused due to cardiovascular infections and its determination is exceptionally troublesome handle. Ordinarily, these maladies can be analyzed utilizing instinct of the therapeutic pro and it would be profoundly advantageous in case the methods utilized for examination might be progressed with the restorative data framework. At diminished taken a toll, in the event that a choice back or computer based data framework is created at that point it'll be supportive for precise determination.

By Utilizing a few information mining procedures heart infection forecast can be made straightforward by utilizing different characteristic to discover out whether the individual endures from heart assault or not, and it moreover takes less time to for the forecast and progress the restorative determination of infections With great exactness and minimizes the event of heart attack [2]. Information mining in conjunction with delicate computing procedures makes a difference to disentangle covered up connections and analyze maladies effectively indeed with instabilities and mistakes.

2. LITRATURE REVIEW

The Cardiovascular Illness Information Forecast is planned to back clinicians in their determination for heart infection expectation. They regularly work through an investigation of restorative information and an information base of clinical skill. The quality of therapeutic demonstrative choices for heart infection can be expanded by changes to these foreseeing frameworks.

Various works has been done related to cardiovascular expectation framework by utilizing different data mining methods and calculations by numerous researchers. The point of all is to realize superior exactness and to create the framework more proficient so that it can foresee the chances of heart assault. This paper points at analysing the different information mining strategies presented in later a long time for heart disease forecast. Distinctive information mining strategies have been utilized within the conclusion of Cardiovascular Disease over diverse CVD datasets.

Raj Kumar et.al.[3] analysed cardiovascular disease using various information mining techniques as Decision Tree, Naïve Bayes and KNN. They utilized a dataset with 3000 tests of information, each test comprise of 14 highlights of characteristics of heart infection. 70% of information is

utilized for instruction and 30% for testing. Performance of these algorithms is measured on the basis of throughput time and time taken.

In another study [4] conducted a determination of cardiovascular malady with KNN. Dataset utilized included 303 tests with 13 highlights of heart patients. In KNN, the K value was relegated from 1 to 13, the most noteworthy exactness was 7 that the proportionate of 97.4% has been gotten.

B. Venkatalakshmi and M.V Shivsankar[5] performed an investigation on cardiovascular disease determination utilizing information mining procedures Naïve Bayes and Decision Tree procedures. Diverse sessions of tests were conducted with the same datasets in WEKA 3.6.0 software. Information set of 294 records with 13 traits was utilized and the comes about uncovered that the Naïve Bayes outflanked the Decision tree procedures.

3. K-NEAREST NEIGHBOURS

The k-nearest neighbours (KNN) calculation may be a basic, easy-to-implement directed machine learning calculation that can be utilized to illuminate both classification and regression issues. The KNN calculation accept that comparative things exist in near vicinity. In other words, comparable things are close to each other.

Steps of KNN algorithm is as given below[6]:

1. Stack the data.
2. Determine K to your chosen number of neighbors.
3. For each illustration within the data

3.1 Calculate the remove between the inquiry case and the current illustration from the data.

3.2 Include the remove and the file of the illustration to a requested collection.

4. Sort the requested collection of separations and records from littlest to biggest (in climbing arrange) by the distances.
5. Choose the primary K passages from the sorted collection.
6. Get the names of the chosen K entries.
7. In case regression, return the cruel of the K labels.
8. On the off chance that classification, return the mode of the K labels.

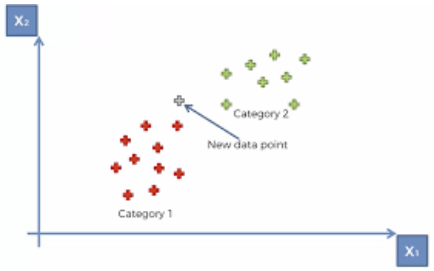

Fig. 1: Dataset Representation for KNN

4. SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) may be a administered machine learning algorithm which can be utilized for both classification or regression challenges. In any case, it is for the most part utilized in classification issues. Within the SVM calculation, we plot each information thing as a point in n-dimensional space (where n is number of features you have) got with the value of each feature being the value of a specific coordinate. At that point, we perform classification by finding the hyper-plane that separates the two classes exceptionally well.

Steps for SVM classifier is as given below:

1. Identify the right hyper-plane in each possible scenario.
2. Classify the classes of data points.
3. Find the hyper-plane to segregate the classes.

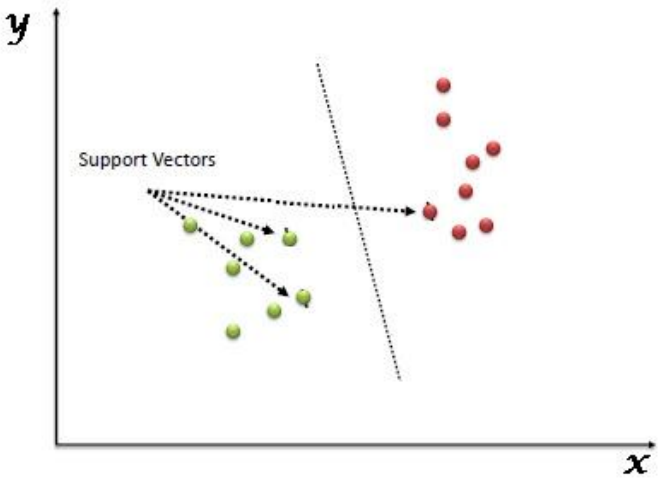

Fig. 2: Dataset representation of Support Vector Machine

5. DECISION TREE

Applying Decision Tree strategies has appeared valuable precision within the determination of heart infection. But assisting health care experts within the conclusion of the world's greatest executioner requests higher precision. Investigations looks for to progress determination precision to move forward health results. Decision Trees (DTs) are a non-parametric administered learning method used for

classification. The most point is to form a show that predicts the value of a target variable by learning straightforward choice rules induced from the data features. The structure of decision tree is within the frame of a tree. Decision trees classify occurrences by beginning at the root of the tree and moving through it until a leaf node. Choice trees are commonly utilized in operations inquire about, basically in decision analysis [7].

The data obtained for the tree is identified using the below given Eq.(1).

$$E(S)= -P(P)\log 2P(P)-P(N)\log 2P(N) \quad (1)$$

Steps for Decision Tree Algorithm is as given:

Step 1: Distinguish the data pick up for the properties within the dataset.
Step 2: Sort the data pick up for the heart malady datasets in plummeting order.
Step 3: After the recognizable proof of the information pick up relegate the finest quality of the dataset at the root of the tree.
Step 4: At that point calculate the data pick up utilizing the same formula.
Step 5: Part the nodes based on the most elevated data pick up value.
Step 6: Rehash the method until each traits are set as leaf hubs in all the branches of the tree.

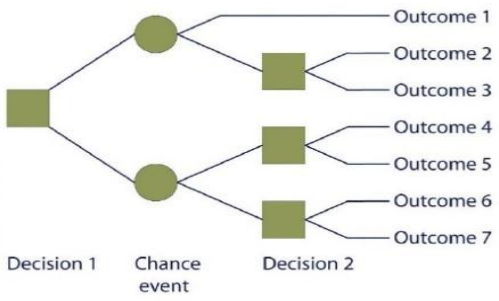

Fig. 3: Decision Tree

6. RANDOM FOREST

Random Forest is basically a gathering of unpruned classification trees. It gives fabulous execution on a number of practical issues, to a great extent since it isn't sensitive to commotion within the information set, and it isn't subject to over fitting. It works quickly, and generally shows a considerable execution change over numerous other tree-based calculations.

There are three fundamental choices to be made when constructing a arbitrary tree. These are:

1. The strategy for part the leaves.
2. The sort of indicator to utilize in each leaf.
3. The strategy for infusing haphazardness into the trees.

Random Forest comprises of decision trees. Each decision tree is formed by subset of training data which arbitrarily chosen. The decision tree may be a strategy for showing a arrangement of laws that are leading to a category or value. The contrast between the strategies of decision tree is that how the distance to be measured. Decision trees that are utilized to foresee the cluster factors called classification trees since they are found the tests in clusters or classes [9].

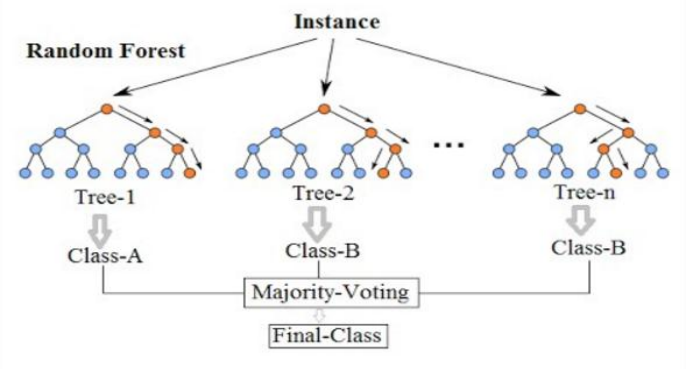

Fig. 4: Random Forest

7. CONCLUSION

On the basis of above studies of various algorithms we reach at the conclusion that for cardiovascular disease detection we can use any one the algorithm i.e. KNN, Support vector Machine. Decision Tree or Random Forest. The performance of these algorithms depends on the dataset, size of dataset, type of disease to be diagnosed. In one of our work to diagnose cardiovascular disease we find the K-Nearest Neighbors algorithm to be more suitable.

REFERENCES

[1] WHOCardiovascularDiseases.http://www.who.int/cardiovascular_diseases.
[2] Neha Chauhan and Nisha Gautam "An Overview of heart disease prediction using data mining techniques".
[3] A. Rajkumar, G.S. Reena, "Diagnosis Of Heart Disease Using Data mining Algorithm", Global Journal of Computer Science and Technology" Vol. 10, Issue 10, Ver. 1.0, pp. 38-43, September 2010.
[4] M. Shouman, T. Turner, R. Stocker, "Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients", International Journal of Information and Education Technology, Vol. 2, No. 3, pp. 220-223, June 2012.
[5] B.Venkatalakshmi, M.V Shivsankar, "Heart Disease Diagnosis Using Predictive Data mining"; International Journal of Innovative Research in Science, Engineering and Technology Volume 3, Special Issue 3, March 2016.

[6] https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761
[7] Prof. Mamta Sharma, Farheen Khan , Vishnupriya Ravichandran, "Comparing Data Mining Techniques Used For Heart Disease Prediction",International Research Journal of Engineering and Technology (IRJET), Volume: 04 Issue: 06 | June -2017
[8] H. Benjamin Fredrick David and S. Antony Belcy, "HEART DISEASE PREDICTION USING DATA MINING TECHNIQUES" ICTACT JOURNAL ON SOFT COMPUTING, OCTOBER 2018, VOLUME: 09, ISSUE: 01.
[9] Serdar AYDIN, Meysam Ahanpanjeh and Sogol Mohabbatiyan, "COMPARISON AND EVALUATION DATA MINING TECHNIQUES IN THE DIAGNOSIS OF HEART DISEASE", International Journal on Computational Science & Applications (IJCSA) Vol.6, No.1, February 2016.

BIOGRAPHIES

Mr. Eqbal Ahmad is a M. tech Scholar in Computer Science and Engineering Department at Bhabha Institute of Technology, Kanpur, and Uttar Pradesh, India. He is in Education Industry from last 13 years, His research areas are Python programming, Data Science, Machine Learning, data structures, DBMS and Artificial Intelligence. He published many papers in various International Journals.

Mr. Arpan Kumar is currently working as an Assistant Professor in Computer Science and Engineering Department at Himalayan Institute of Technology and Management, Lucknow, Uttar Pradesh, India. His research areas are Digital Image Processing, Data Mining, Neural Network and Artificial Intelligence. He is lifetime member of various scientific and researchers' societies. He published many papers in various International Journals.