

# Exploration Of A State Of The Art On Cardiac Diseases Prediction Techniques

S.Usha<sup>1</sup>, Dr.S.Kanchana<sup>2</sup>

<sup>1</sup>Research Scholar, Department of Computer Science,  
SRM Institute of Science & Technology, Kattankulathur, Chennai.

<sup>2</sup>Assistant Professor, Department of Computer Science,  
SRM Institute of Science & Technology, Kattankulathur, Chennai.

## ABSTRACT –

*Healthcare is a predictable task to wipe out human life. Coronary heart disease is sickness that impacts the human coronary heart. Cardiovascular sicknesses will forecast with the aid of several techniques that helped in making choices about the modifications that maintain excessive-risk patients which resulted in the discount of their dangers. The purpose of demise ratio of those sicknesses may be very high. It is very imperative to become aware of if the individual has heart disorder or now not. In medical field it is very important to find the occurrence of prediction of the heart diseases. Accurate Prediction results are very efficient to treat the patient's medical history before the attack occurs. The techniques Data mining and Machine learning plays a essential role to predict the occurrence of heart diseases. These techniques diagnose these diseases with the help of dataset in healthcare centers. Various models used to reduce the number of deaths ratio. Models based on several algorithms such as Support Vector Machine (SVM), Decision Tree(DT), Naïve Bayes(NB), K-Nearest Neighbor(KNN), and Artificial Neural Network (ANN) are implemented to predict heart disease. The accuracy of these models helps to diagnose the diseases with better results. This paper summarized the performance of all algorithms which are used to predict and diagnose heart diseases.*

**KeyWords:** Artificial Neural Network, Data Mining, Decision Tree, K-Nearest Neighbor, Machine Learning.

## 1. INTRODUCTION

### A. Heart Disease

Heart is considered as the major part of the human being body. All parts of human body received blood from heart because it circulate and pump blood to all the parts. Human life depends on the heart. If blood circulation from heart to brain is stopped then all neuron nerve systems will not work. This may occur to death. A healthy life is intimate that the heart is functioning properly. A disorder that disturbs the function of the heart is called heart disease. The most prominent causes of death all over the world are heart disease. It is a very challenging and more complicated task to predicate heart disease. It is challenging work for health care centers to analyses the disease from massive amounts of data. The different kind of heart disease are Coronary artery disease (CAD), Cardiomyopathy, Hypertensive heart disease, Inflammatory heart disease, Valve heart disease and failure (HF). In medicine, diagnosis of diseases is a very tedious task. The cardiovascular disorder commonly refers to situations that contain narrowed or obstructed blood vessels that may cause chest pain (angina), a heart assault, and stroke. Other cardiac situations, like those who have an effect on

your coronary heart's muscle, valves or rhythm, are also recognizable forms of coronary heart disease.

A heart assault happens whilst the blood drift to heart that obstructed through a blood clot. The blood go with the flow perfectly, when the clot is cut off. The part of the heart muscle provided through that artery starts off evolved to die. The prediction of coronary heart ailment is analyzed by using traditional techniques. Data Mining and machine mastering are both strategies used to diagnose heart illnesses

### *B. Data Mining*

Extracting data from large databases said to be Data mining. Data mining is also called as the Knowledge discovery process, knowledge mining data. It includes data collection, data analysis, and data extraction. Data Mining is a process to find out the significant information to look at meaningful data. Various fields use data mining techniques to predict their problems. One of the important fields is Health Care. Health care plays a vital role in using Data Mining techniques.

Data mining techniques are used to detect unknown and valuable information in health data. It detects the fraud and causes of diseases in health insurance, provides low cost to the patients for the availability of medical solutions and identifies various methods for the medical issues. It is very difficult to analyze the data and to make decisions for health organizations regarding patient health when the medical data is vast and complex. Analysis of data will enhance the performance of patient management tasks. [1] Data mining techniques can be deployed in the automated system. This will help the patients to reduce the test for their disease.

### *C. Machine Learning*

Machine Learning has an impact full in the Healthcare industry. Some of its applications in health care areas include medical diagnosis. Machine Learning is defined as a hard and fast of methods that identify the styles routinely and used to discover the information that predict destiny records to perform a few form of choice making. Machine learning tells computer to perform tasks by providing some examples[2].

Machine learning types Supervised Learning, Unsupervised Learning and Reinforcement Learning used to train labeled and unlabeled data. Classification and regression are two common supervised learning types. Clustering is belongs to unsupervised types. Most clinical problems in the medical field were solved using machine learning methods. The method will predicate and diagnose all diseases that help to find the accurate result which helps the patients in the medical field. Many researchers contribute their work to improve the development of the medical field. It is a very curious case for researchers to predict diseases in this field. Data mining and Machine Learning are techniques that are very useful for the researcher to focus on their work. This paper summarized the state of art work by many researchers.

## **2. TECHNIQUES**

### *1. Classification*

Classification is part of supervised learning techniques, which is employ widely in various fields. It predicts a class for both input and output variables. It also states that the data elements which belong to class gives the best result when the output is finite and discrete values. [3] Suggested that the group membership is predicted by classification techniques for data instances. A classification technique includes Decision tree induction, Bayesian

networks, K-Nearest Neighbor classifier to predict the model. [16] This technique proves an efficient result in diagnosis of heart diseases.

## 2. Decision Tree

Decision tree consists of nodes, branches and leaf nodes. The structure of decision tree is like a tree format. Every node in a tree represents a condition on the input attributes. Branches from each node describe the outcome of the condition. Classification is occur in the leaf node which holds a class label and assigns dataset record according to those classes.

The continuous variables cannot handle directly Decision Tree is a data mining techniques that not able to handle continuous variables directly. So the continuous attributes must be converted to discrete attributes [4]. Decision Tree use binary discretization for continuous-valued features. Error pruning method is applied to decision tree to improve the accuracy. Systematically investigates accuracy by comparing the multiple classifiers [4]

Decision tree is representing in the form tree to take specific decision in all situation that used for predictive model. Root, nodes and branching are main component of decision tree. Many procedures such as ID3, CART, CYT, C5.0 and J48 of decision tree have used to classify the dataset.

A decision tree is belongs to supervised gaining knowledge of strategies in Machine Learning algorithms. It's miles used for both type and regression. From this algorithm, statistics may be break up in step with the attributes. A decision tree is a tree in order that include nodes and leaves. At leaves, we will get effects or selection, and at the nodes, data could be split. Decision tree is used in those areas of the medical science where numerous parameters involved in classification of data set.

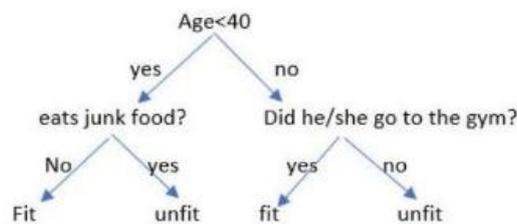


Fig 1. Decision Tree Example

## 3. Naive Bayes

A Naive Bayes classifier also called as an independent feature model is a simple probabilistic classifier that depends on Bayes theorem with strong i.e. naive independence assumptions. In Naive Bayes classifier the presence (or absence) of a selected feature of a category is unrelated to the presence (or absence) of another feature. Naive Bayes classifiers are considered as a type of supervised learning. A Naïve Bayesian model is simple to build and it is particularly useful in the field of clinical technological know-how for diagnosing heart sufferers. Earlier it's far used for text class. It deals with the datasets which have the very best dimensionality. a few examples are sentimental evaluation, spam filt ration, and so on. This naïve Bayes algorithm is based on Bayes theorem with the assumption that attributes are unbiased of each different

## 4. Support Vector Machines

Support Vector Machines (SVMs, also support vector networks) are Labeled learning models that evaluate data to make an output in a patterns. In this algorithm a labeled training data will make a classifier that divide the particular data into several classes.

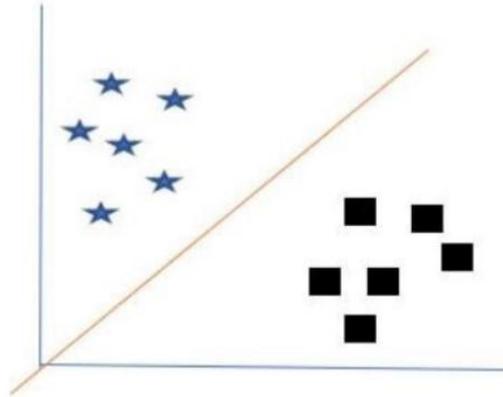


Fig 2. SVM Example.

### 5. *Neural Network*

An artificial neural network is computational networks which project to indicate the networks of neuron (neuron) of the organic critical machine a apprehensive. The structure of a neural network is made via variety of processing units (neurons) and connection between them. Each connection have weights which is associated with them to represent their strength.

A subgroup of processing factors is known as a layer within the network. It involves 3 exclusive layers. The number one layer is that the input layer which represents an interface with the surroundings and therefore the ultimate layer is that the output layer in which output is saved. Between the input and output layer there can also be one and greater hidden layers .there's no obstacle on determining the quantity of hidden layer. The quantity of neurons at some stage in a layer and therefore the number of layers relies upon on the difficulty of the device studied. Consequently, the top-quality specification needs to be determined.

### 6. *Random Forest*

Random forest version constructs quantity of choice timber and locate mode of all training output through person tree as a final output. There can be high variance in unmarried tree however RF lightens the hassle of excessive variance. Random forest changed into built by using schooling 303 times of data and authentication of accuracy was accomplished the usage of 10- fold move validation[6]. This set of rules allows many sufferers to treatment from the coronary heart disorder. The technique of predicting coronary heart illnesses the use of Random wooded area with properly-set attributes bring us greater accuracy.

## 3.LITERATURE SURVEY

The presence and absence of disease is distinguish with the help of Feed-forward back propagation algorithm. This algorithm is used to classify the disease. The hidden layer of neurons is selected using test and error method. This algorithm produced 88% of result to classify the data [7].

In KNN classifier more number of predominant attributes is used to achieve approximately 80% of accuracy [8]

Optimized SVM improve the accuracy of prediction with the help of k-fold cross validation. These use to diagnosis diseases at an early stage and reduce the cost. The result obtained is better way to predict the disease [9].

In health organization implementation of data classification and extraction of knowledge is described by different classifier. With help of this classifier it came to know that when and how to use [10]

Various Data mining techniques are applied to analyze the prediction of heart disease using vital parameter to extract the pattern from the dataset [11].

So as to reduce range of deaths from heart diseases decision tree is used. This technique is very efficient and quick to detect the result. Different algorithms of decision tree classification are compared to get the better performance. The algorithms are J48 algorithm, Logistic Model Tree and Random Forest algorithm [12]. Main goal is to extract hidden pattern to diagnosis heart disease[12].

Two separate datasets have used for diagnosis heart disease. All patients' data are trained by different classifier algorithm. The accuracy rate given by SVM and Naïve Bayes algorithm is 95.5% and 73.5%[13].

The classification uses the Decision Tree and K-Nearest Neighbor (KNN) Algorithms as supervised classification model to calculate and compare the algorithm used to diagnosis. The Pima Indians Dataset is applied in this system for research[14].

Risk model is proposed to predict heart failure, from medical reports. Multinomial Naïve Bayes is used to perform multi-class classification.. The data which was taken from the Armed institute of Cardiology are based and unstructured layout. 86.7% and 92.4%, of accuracy and Area under the Curve (AUC) is achieved by this model [15]

To get accuracy 85.81% Random Forest was applied to exploit the non-linear heart disease dataset [6]. Challenged was taken up to increase the accuracy of prediction.

To envisage heart diseases DNFS technique is used which is based on fuzzy neural system. This system is predicted with the aid of data mining and machine learning techniques. The proposed system produces best accuracy after comparing the algorithm [17].

#### 4. CONCLUSION & FUTURE WORK

This paper summarized the hassle using numerous algorithms of Data mining and Machine learning in the subject of clinical discipline. The primary consciousness is to mix numerous characteristic for sensible and effective coronary heart attack. The outcome of predictive facts processing method on an equal dataset well-known shows that name Tree outperforms and a some occasion Bayesian classification has comparable exactness as of selection tree but other extrapolative strategies like KNN, Neural Networks, category supported clustering are not executed properly. Next end is the accuracy of the Decision Tree and Bayesian class additional improves after using genetic algorithm to scale back the scrupulous information bulk to induce the top-quality subset of characteristic sufficient for coronary heart circumstance prediction. The current work is frequently further greater and improved for the automation of coronary heart condition prediction. Actual facts from health care businesses and companies must be collected and each one the available techniques are going to be as compared for the most beneficial accuracy.

#### 5. REFERENCES

- [1] N. Bhatla and K. Jyoti, "An Analysis of Heart Disease Prediction using Different Data Mining Techniques." [Online]. Available: [www.ijert.org](http://www.ijert.org).
- [2] B. Dhomse Kanchan and M. Mahale Kishor, "Study of machine learning algorithms for special disease prediction using principal of component analysis," in *Proceedings - International Conference on Global Trends in Signal Processing, Information Computing and Communication, ICGTSPICC 2016*, Jun. 2017, pp. 5–10, doi: 10.1109/ICGTSPICC.2016.7955260.
- [3] G. Kesavaraj and S. Sukumaran, "A study on classification techniques in data mining,"

- 2013, doi: 10.1109/ICCCNT.2013.6726842.
- [4] C. R. Olsen, R. J. Mentz, K. J. Anstrom, D. Page, and P. A. Patel, "Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure: Machine learning in heart failure," *American Heart Journal*, vol. 229. Mosby Inc., pp. 1–17, Nov. 01, 2020, doi: 10.1016/j.ahj.2020.07.009.
- [5] M. A. jabbar, B. L. Deekshatulu, and P. Chandra, "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm," *Procedia Technol.*, vol. 10, pp. 85–94, 2013, doi: 10.1016/j.protcy.2013.12.340.
- [6] Y. K. Singh, N. Sinha, and S. K. Singh, "Heart disease prediction system using random forest," *Commun. Comput. Inf. Sci.*, vol. 721, pp. 613–623, 2017, doi: 10.1007/978-981-10-5427-3\_63.
- [7] N. Ajam, "Heart Diseases Diagnoses using Artificial Neural Network," *Netw. Complex Syst.*, vol. 5, no. 4, pp. 7–11, 2015, [Online]. Available: www.iiste.org.
- [8] N. Khateeb and M. Usman, "Efficient heart disease prediction system using K-nearest neighbor classification technique," in *ACM International Conference Proceeding Series*, Dec. 2017, pp. 21–26, doi: 10.1145/3175684.3175703.
- [9] A. Davari Dolatabadi, S. E. Z. Khadem, and B. M. Asl, "Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM," *Comput. Methods Programs Biomed.*, vol. 138, pp. 117–126, 2017, doi: 10.1016/j.cmpb.2016.10.011.
- [10] M. Gandhi and S. N. Singh, "Predictions in heart disease using techniques of data mining," in *2015 1st International Conference on Futuristic Trends in Computational Analysis and Knowledge Management, ABLAZE 2015*, Jul. 2015, pp. 520–525, doi: 10.1109/ABLAZE.2015.7154917.
- [11] R. Chadha and S. Mayank, "Prediction of heart disease using data mining techniques," *CSI Trans. ICT*, vol. 4, no. 2–4, pp. 193–198, 2016, doi: 10.1007/s40012-016-0121-0.
- [12] S. B. Patel, "Heart Disease Prediction Using Machine Learning and Data Mining Technique," no. October, 2016, doi: 10.090592/IJCSC.2016.018.
- [13] K. Deepika and S. Seema, "Predictive analytics to prevent and control chronic diseases," *Proc. 2016 2nd Int. Conf. Appl. Theor. Comput. Commun. Technol. iCATccT 2016*, pp. 381–386, 2017, doi: 10.1109/ICATCCT.2016.7912028.
- [14] E. K. Hashi, M. S. Uz Zaman, and M. R. Hasan, "An expert clinical decision support system to predict disease using classification techniques," in *ECCE 2017 - International Conference on Electrical, Computer and Communication Engineering*, Apr. 2017, pp. 396–400, doi: 10.1109/ECACE.2017.7912937.
- [15] M. Saqlain, W. Hussain, N. A. Saqib, and M. A. Khan, "Identification of Heart Failure by Using Unstructured Data of Cardiac Patients," *Proc. Int. Conf. Parallel Process. Work.*, vol. 2016-September, pp. 426–431, 2016, doi: 10.1109/ICPPW.2016.66.
- [16] K. Polaraju, D. Durga Prasad, "Prediction of Heart Disease using Multiple Linear Regression Model", *International Journal of Engineering Development and Research Development*, ISSN:2321- 9939, 2017.
- [17] "Analysis and Prediction of Various Heart Diseases Using Dnfs Techniques," vol. 2, no. 1, pp. 1–7, 2015. 978-1-7281-4108-4/20/\$31.00