iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** VII **Month of publication:** July 2020

DOI: http://doi.org/10.22214/ijraset.2020.7019

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VII July 2020- Available at www.ijraset.com

A Survey on Heart Disease Prediction using various Machine Learning Techniques

Vinutha H M[1], Dr. K Thippeswamy[2]
[1]*Department of Computer Science and Engineering, VTU CPGS, Mysore*
[2]*Department of Computer Science, VTU CPGS, Mysore*

Abstract: ***Heart disease is most life-threatening disease in the world today. Diagnosing a patient with heart disease has turned into a challenging work in the area of medical science. Huge amount of cardiovascular health study dataset collected from the healthcare industries can be used to predict heart disease. In this work we compare different machine learning techniques such as Naïve Bayes, Support Vector Machine (SVM), Decision Tree, K-Nearest Neighbor (KNN), Random Forest and Artificial Neural Network (ANN). This paper provides the overall information about the performance and accuracy of various machine learning technique in heart disease prediction.***
Keywords: Machine learning, Decision Tree, K Nearest Neighbor (KNN), Naïve Bayes, Artificial Neural Network (ANN).

I. INTRODUCTION

Heart play important role in circulatory system as it pumps blood and because of it other body parts get enough amount of oxygen and nutrients. When a heart disease occurs, it cost huge medical care and permanent impairment but can eventually lead to death. If we were able to predict heart disease in early stage we can save unwanted medical tests and diagnose the patient in a right way. We can save patient life if heart disease is predicted in an early stage for this reason many healthcare research are going fast to reduce death rates. Huge amount of cardiovascular healthcare datasets collected from hospitals contains some hidden information that need to be extracted with the help of data mining techniques. The main aim of any disease prediction system is to take a medical decision which is a difficult and challenging job. Different machine learning classification methods such as Naïve Bayes, support vector machine, k nearest neighbor, decision tree, Random Forest and Artificial Neural Network, will take medical dataset and produces set of appropriate prediction. The above techniques can extract hidden knowledge from huge amount of clinical dataset and can predict patient with disease outcome based on the risk factors like age, blood pressure, pulse rate, cholesterol, blood sugar, etc. These machine learning techniques have the ability to answer complex queries and each of it has its own strength and accuracy.

The advantage of using these machine learning techniques is that these can predict the disease in less amount of time and less amount of cost. Hence reducing extra work need to be performed by the medical practitioners. To improvise the system for effective prediction more research need to be done in the area of machine learning. In order to predict the disease we need to classify the datasets into many distinct classes.

Classification is technique to categorize datasets into a desired and distinct number of classes where we can assign label to each class. We use different types of algorithms for classification in machine learning some of them are mentioned below.

1) *Naive Bayes:* Naive Bayes is a classification technique based on the Bayes theorem, we can assign class labels to each features based on the probability calculated by the algorithm. It assumes that all the values in the particular features are independent to each other.
2) *Support Vector Machine:* SVM is a non-probabilistic binary linear classifier used to solve both classification and regression problems. It allows us to classify linearly separable data. The data points present in the hyper plane it makes separation of those using decision boundary.
3) *K-Nearest Neighbor (KNN):* The k-NN algorithm assumes instances in the n dimensional space. The distance between instances are calculated using standard Euclidean distance. The instances with a smaller distance are classified as one group in an instance-based classifier.
4) *Decision Tree:* A decision tree is a classification technique used to construct a tree like structure using the data set which it contains root node internal node, branches and leaf nodes, where each branches represents some course of action and every node is the outcome of that action or decision. The classification rules can be represented as a paths from root node to leaf node.
5) *Random Forest:* As name suggests it contain multiple connected decision tree. It select's N random records from dataset based on these records it will build individual decision tree.

6) *Artificial Neural Network:* This algorithm is used mainly for the classification it contain processing units called neuron which takes the dataset as input and produces the output. It is a multilayered structure where it consist of input layer hidden layer and output layer.

II. LITERATURE SURVEY

In the literature survey we compare different classification algorithms used to detect the possibility of occurring heart disease

Anjan Nikhil Repaka et al. [4] developed the heat disease prediction system using Naive Bayes classifier. The main attributes used for prediction are age, blood pressure, cholesterol, sugar, gender. These attributes are given as input to the navies' byes classifier. The data collected from the UCI repository where it consist of 297 patient records 80% of the data is used for testing and 20% is used for training. The naive Bayes algorithm is performed in supervised learning. In navies byes the value of the attributes available in one class is independent of the value of other class.in addition the author has provided the security for the data using Advanced Encryption System (AES). The result showed that we can get 89% of accuracy by using naive Bayes classifier.

Mr.Santhana Krishnan.J et al. [7] developed a prototype that predicts out coming possibilities of heart disease. The output of the system shows possibility of occurring heart disease in terms of percentage by applying decision tree and Naive Bayes classifier algorithms to the datasets and compared both of the classification algorithm to same set of data. The dataset are collected from UCI repository which contains 300 instances and 14 attributes. Weka tool is used for classification, 70% of the data is used for training and 30 % of the data is used testing. The Naive Bayes classifier algorithm predicted the disease with 87% of accuracy and the Decision tree classifier algorithm predicted the disease with 91% of accuracy.

Divya Krishnani et al. [5] developed system that can predict heart disease using "Framingham Heart Study" dataset, which contains 4240 number of patient records with 16 attributes. The system aims at predicting only one type of heart disease that is coronary heart disease. The system compares the performance of three classification algorithms namely Decision tree, Random Forest and k nearest neighbor. Among 4240 number of instances only 644 number of patients are suffers from coronary heart disease therefore author have extracted only that amount of records. To replicate the instances and balance the class distribution the author have used random over sampling method. The result showed that decision tree, random forest and k nearest neighbor can predict the disease with 92.7%, 96.8% and 92.8% of accuracy respectively.

Purushottam et al. [2] proposed a framework for medical diagnosis that improve medical care and it can also decrease costs for unwanted clinical tests. Based on the clinical attributes like age, blood pressure, blood sugar the author have implemented system that can identify the risk level of a patient with heart disease. The datasets are extracted from Cleveland database with 303 where 50% of the data is used for training and remaining 50% is used for testing. These attributes can be prioritized based on the requirement of a user. The performance of the classification algorithm is evaluated in terms of accuracy. The classification algorithm used is decision tree. The outcome of the framework has great ability in predicting the risk level of heart disease with accuracy 86.7%.

Thankgod Obasi M et al. [8] proposed a model that can predict heart disease in human using existing medical record of patient as dataset for testing and training. The total number of instances used were 4832, 80% of the data is used for the training and 20% of the data is used for testing. The classification algorithms used to classify the datasets are Random Forest, Naive Bayes and Logistic Regression and these classification algorithms can predict the disease with accuracy92.4%, 61.9%, and 59.7% respectively.

Ritika Chadha et al. [3] developed the heart disease prediction using the classification methods like ANN, naïve bayes, and decision tree they have fed the datasets to all the three classification model each model showed 100% accuracy for artificial neural network method, 85% accuracy for decision tree and 88% accuracy for naïve bayes. They have used WEKA tool and python for building the classification model

Kapil Wankhade et al. [1] proposed a decision support system for predicting heart disease using two classification algorithms namely artificial neural network (ANN) and support vector a machine. Database is extracted from UCI repository 13 attributes are used for classification. A multilayer perceptron with backpropogation is used to train the datasets. The result shows that support vector machine can classify the disease with accuracy 80% and artificial neural network can classify and predict the heart disease with accuracy 98%.

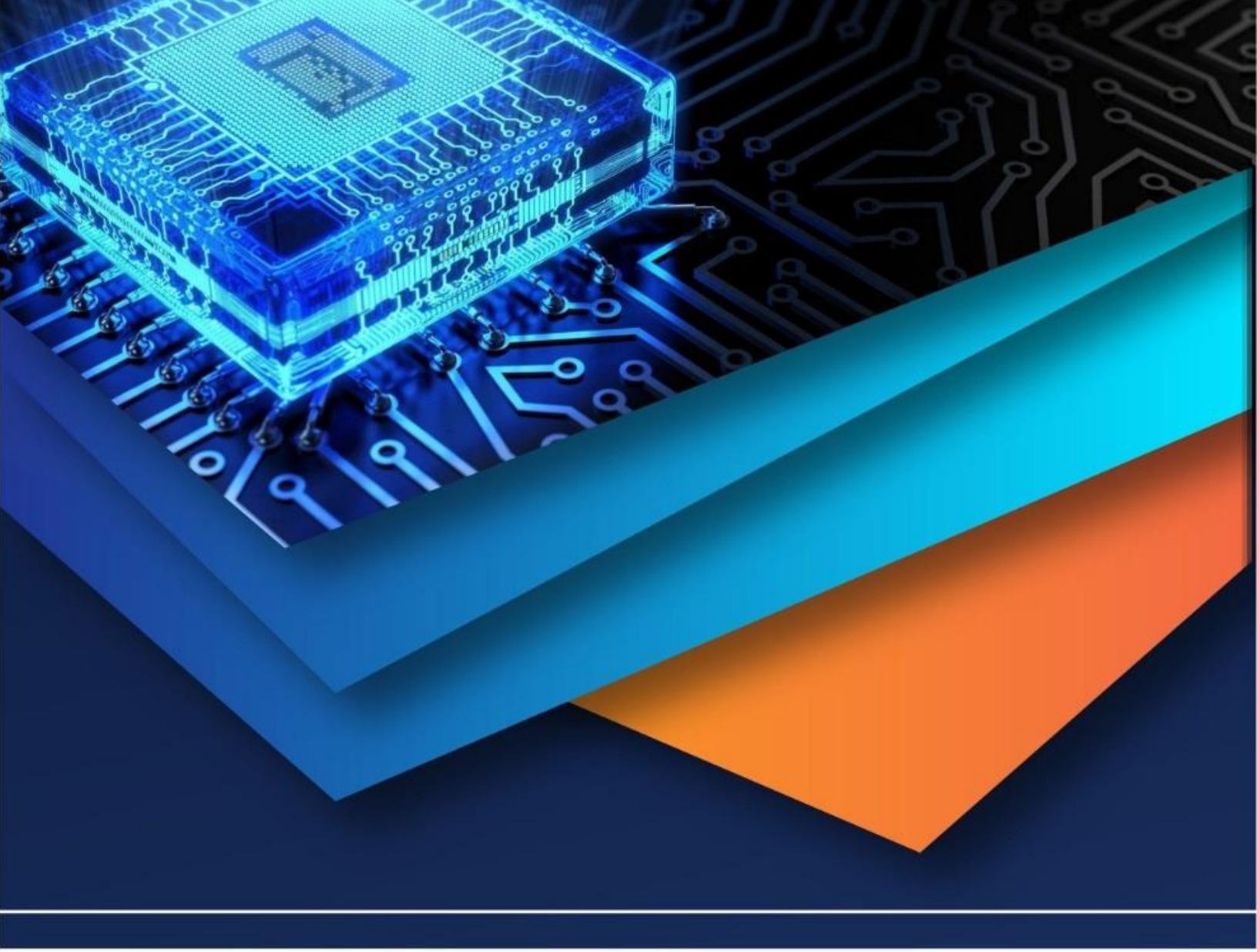
G. Suseendran et al. [6] designed a heart disease prediction model using principal component analysis (PCA), logical binary pattern (LBP) and artificial neural network, where LBP is used for feature extraction and PCA is used for feature set size reduction and neural network with backpropogation is used for classifying datasets. The result showed that artificial neural network classification technique can predict the disease with accuracy 95%.

III. CONCLUISON

This paper surveys about various classification techniques used to classify the datasets and predict the heart disease based on the hidden patterns present in that database. The prediction of the heart disease can be done in less amount and time and cost using machine learning algorithms. After doing a clear study on all classification algorithms it shows that we can get higher accuracy in disease prediction using artificial neural network technique. Therefore we can come to a conclusion that artificial neural network classification technique can classify and predict the heart disease with higher accuracy compared to other classification algorithms.

REFERENCES

[1] Kapil Wankhade,Mrudula Gudadhe, and Snehlata Dongre, Sept 2010."Decision Support System for Heart Disease Based on Support Vector Machine and Artificial Neural Network".
[2] Purushottam, Kanak Saxena, Richa Sharma"Efficient heart disease prediction system using Decision tree" 2015.
[3] Ritika Chadha1, Shubhankar Mayank "Prediction of heart disease using data mining techniques"2016.
[4] Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin" Design And Implementing Heart Disease PredictionUsing Naives Bayesian" 2019.
[5] Divya Krishnani, Anjali Kumari, Akash Dewangan, Aditya Singh, Nenavath Srinivas Naik "Prediction of Coronary Heart Disease usingSupervised Machine Learning Algorithms" 2019.
[6] G. Suseendran, Noor Zaman, M. Thyagaraj, R. K. Bathla" Heart Disease Prediction and Analysis using PCO, LBP and Neural Networks" 2019.
[7] Santhana Krishnan.J, Dr.Geetha.S" Prediction of Heart Disease Using Machine Learning Algorithms."2019.
[8] Thankgod Obasi, M. Omair Shafiq," Towards comparing and using Machine Learning techniques for detecting and predicting Heart Attack and Diseases" 2019.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY