Heart Disease Prediction based on Machine Learning and Deep Learning Techniques

Supriya Kaskar[1],
Department of Computer Engineering,
ICOER, Wagholi, Pune

Sonali Gandewar[2],
Department of Computer Engineering,
ICOER, Wagholi, Pune

Sapna Boinwad[3]
Department of Computer Engineering
ICOER, Wagholi, Pune

Pradnya Bhusare[4]
Department of Computer Engineering
ICOER, Wagholi, Pune

Dr. Vinod Wadne[5]
Department of Computer Engineering
ICOER, Wagholi, Pune

---***---

Abstract: Heart disease is one of all the foremost vital issues that are arising within the world these days. disorder prediction could be a crucial challenge within the space of clinical knowledge analysis. Hybrid Machine learning (ML) has been showing an efficient help in creating selections and predictions from the massive amount of information created by the care industries and hospitals. We've got additionally seen mil techniques being employed in recent developments in several areas of the net of Things (IoT). Varied studies provide solely a glimpse in predicting cardiovascular disease with mil techniques. During this paper, we have a tendency to propose a narrative methodology that aims at finding vital options by applying machine learning techniques that lead to up the accuracy within the prediction of disorder. The prediction model is projected with mixtures of various options and a number of other classification techniques. We have a tendency to turn out associate degree increased performance level with an associate degree accuracy level of ninety-two through the prediction model for cardiovascular disease with the hybrid random forest with a linear model.

Keywords: Machine Learning, Supervised Learning, Heart disease, prediction, Imbalance data, etc.

INTRODUCTION

In today's world of stress Heart, being an important organ during a soma that pumps blood through the body for the blood circulation is crucial and its health is to be preserved for a healthy living. The health of somebody's heart relies on the experiences during a person's life and is totally captivated with skilled and private behaviours of an individual.

There may additionally be many genetic factors through that a kind of heart condition is passed down from generations. in line with the globe Health Organization, per annum quite twelve million deaths square measure occurring worldwide thanks to the varied forms of heart diseases that is additionally renowned by the term upset. The term heart condition includes several diseases that square measure numerous and specifically have an effect on the center and also the arteries of somebody's being. Even young aged individuals around their 20-30 years of period are becoming stricken by heart diseases.

The increase within the chance of heart condition among young is also thanks to the dangerous consumption habits, lack of sleep, restless nature, depression and various alternative factors like blubber, poor diet, case history, high pressure level, high blood cholesterol, idle behavior, case history, smoking and high blood pressure.

The identification of the center diseases may be a important and is itself the foremost sophisticated task within the medical field. All the mentioned factors square measure taken into thought once analyzing and understanding the patients by the doctor, through manual check-ups at regular intervals of your time, the symptoms of heart condition greatly rely on that of the discomfort felt by a personal. Some symptoms don't seem to be sometimes known by the folk. However, common symptoms embrace hurting, dyspnoea, and heart palpitations.

The hurting common to several forms of heart condition is thought as angina, or heart disease, and happens once a district of the center doesn't receive enough elements. Angina is also triggered by disagreeable events or works up and ordinarily lasts below ten minutes. Heart attacks can even occur as results of differing types of heart condition. The signs of a heart failure square measure like angina except that they'll occur throughout rest and have a tendency to be a lot of severe.

The symptoms of a heart failure will typically stomach upset. Abdomen ache will occur, yet as a significant feeling within the chest. alternative symptoms of a heart failure embrace pain that travels through the body, for instance from the chest to the arms, neck, back, abdomen, or jaw, light-weight headedness and dizzy sensations, lush sweating, nausea and expulsion.

Heart failure is additionally AN outcome of heart condition, and dyspnoea will occur once the center becomes too weak to flow into blood. Some heart conditions occur with no symptoms in the least, particularly in older adults and people with polygenic disorder.

LITERATURE SURVEY

[1According to Ordonez the center sickness may be foreseen with some basic attributes taken from the patient and in their work have introduced a system that features the characteristics of a private individual supported all thirteen basic attributes like sex, vital sign, steroid alcohol et al. to predict the probability of a patient obtaining tormented by heart condition. they need accessorial 2 a lot of attributes i.e. fat and smoking behavior and extended the analysis dataset, and not handled massive dataset[2] Yilmaz, have planned a way that uses method of least squares support vector machine (LS-SVM) utilizing a binary call tree for classification of cardiotocogram to seek out out the patient condition. Duff, et al. have done a look work involving 5 hundred and xxxiii patients United Nations agency had suffered from pathology and that they were integrated within the analysis of heart condition chances, major disadvantages of this paper has Underfitting drawback and not handle noise and outliers with efficiency Frawley, et al. have performed a piece on prediction of survival of Coronary heart condition (CHD) that may be a difficult analysis drawback for medical society. They conjointly used 10-fold cross-validation ways to work out the impartial estimate of the 3 prediction models for performance comparison functions. the prevailing system having variety of disadvantages accuracy of the formula is extremely low and not meet the performance. [4]In the prevailing system mining techniques can-not handle economical higher cognitive process result. To little size knowledge set square measure used for prediction and it's have an effect on to the performance of the system and model cannot meet the quantify ability of the formula. Nowadays, heart condition prediction has been a significant construct in the recent world that's impacting society towards health. [5] the most construct is to spot

the people and pulse rate victimization the Random forest formula. Our project tells, however, the center rate and the condition is calculably supported the inputs like vital sign and plenty of a lot of being provided by the user to a system.[6] heart condition has created loads of significant involvement among researchers; one in all the foremost challenges in heart condition is correct detection and finding the presence of it within a person's. Early techniques haven't been most economical find it even medical prof aren't most economical enough in predicting the center sickness [7] data processing consolidates measurable investigation, AI, and info technology. data processing has been connected in a very few regions of restorative administrations such as revealing connections among analysis info and place away clinical info. the leading-edge healthful conclusion may be a composite procedure that needs actual patient knowledge, several long stretches of clinical expertise, and good info of restorative writing.[8] several patients of these days starting from toddlers to senior individuals visit hospitals and clinics with considerations and complains of chest pains and pains with reference to each the correct and left chest. With them having no clue what the difficulty is, this is often wherever sickness prediction comes into the image and filtering it to heart condition prediction by assessing the regions of pain and conducting some important tests. The results from the check just like the steroid alcohol level, bp, and also the cardiogram isn't enough to precisely predict what or disease the patient has

BACKGROUND

The most important background of machine learning algorithms their technique and mathematical formulation are outlined in this section. Analysing the Healthcare data

1. Machine Learning

Machine learning algorithm can be group into two main categories, they include

1. **Supervised Learning**: supervised learning algorithm main feature is target variable and outcome variable to predict. Supervised learning technique is achieved using regression and classification problem.

2. **Unsupervised learning**: in unsupervised learning algorithm no target or outcome variable to predict. It is used for clustering entities into an different groups.

2. Classification Algorithms:

Classification algorithms work by predicting the simplest cluster to that a knowledge purpose belongs to by learning from labelled observations; it uses a group of input options for the educational method. Classification algorithms square measure sensible for grouping knowledge that square measure ne'er seen before into their numerous groupings and square measure thus extensively employed in machine learning tasks.

3. Evaluation Matrix:

1. Accuracy:

it is measured how many true positive and true negative cases is correct. Mathematically it is defined as

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$$

2. Sensitivity or Recall:

1. Recall: tells us how many of the actual positive cases we were able to predict correctly with our model. Mathematically it is defined as

$$\text{Recall} = \frac{TP}{TP+FN}$$

2. Specificity: tell us how many times classifier gets true negative correct value, mathematically is defined as

$$\text{Specificity} = \frac{TN}{TN+FP}$$

3. Precision:

Precision tells us how many of the correctly predicted cases actually turned out to be positive. Mathematically it is defined as,

$$\text{Precision} = \frac{TP}{TP+FP}$$

PROPOSED SYSTEM

After evaluating the results from the existing methodologies, we have used python and pandas operations to perform heart disease classification for the data obtained from the UCI repository. It provides an easy-to-use visual representation of the dataset, working environment, and building the predictive analytics. ML process starts from a pre-processing data phase followed by feature selection based on data cleaning, classification of modeling performance evaluation. The random forest technique is used to improve the accuracy of the result.

1. Increased accuracy for effective heart disease diagnosis.

2. Handles roughest (enormous) amount of data using random forest algorithm and feature selection.

3. Reduce the time complexity of doctors.

4. Cost-effective for patients.

IMPLEMENTATION

Data Preprocessing:

Heart disease information is pre-processed by exploiting varied collections of records. The dataset contains a complete of 303 patient records, wherever half-dozen records square measure with some missing values. Those half-dozen records are aloof from the dataset and also the remaining 297 patient records square measure utilized in pre-processing

Feature Selection:

Among the thirteen attributes of the info set, 2 attributes concerning age and sex square measure wont to determine the private data of the patient. The remaining attributes square measure thought of necessary as they contain important clinical records. Clinical records square a measure important to the identification and learning the severity of heart condition.

Classification Modelling:

The clump of datasets is completed on the premise of the variables and criteria of call Tree (DT) options. Then, the classifiers square measure is applied to every clustered dataset so as to estimate its performance. the most effective acting models square measure known from the on top of results supported their low rate of error.

Decision Tree

For coaching samples of information D, the trees square measure created supported entropy inputs. These trees square measure merely created in a very high down the algorithmic divide and conquer (DAC) approach. Tree pruning is performed to get rid of the irrelevant samples on D. Entropy = −Xm j=1 terrorist group log2 terrorist group

Algorithm for Decision Tree-Based Partition Require:

- o Input: D dataset – features with a target class for ∀features do
- o for Each sample
- o do Execute the Decision Tree algorithm end for Identify the feature space f1, f2, . . . , fx of dataset.
- o end for Obtain the total number of leaf nodes l1,l2,l3,. . . ,
- o ln with its constraints Split the dataset D into d1, d2, d3,. . . , dn
- o based on the leaf nodes constraints. Output: Partition datasets d1, d2, d3,

Random Forest:

This ensemble classifier builds several decision trees and incorporates them to get the best result.

For tree learning, it mainly applies bootstrap aggregating or bagging.

For a given data, X = {x1, x2, x3, ..., xn} with responses Y = {x1, x2, x3, ..., xn} which repeats the bagging from b = 1 to B

BLOCK DIAGRAM

The proposed module can be divided in to different sections, machine learning, Flask, HTML, CSS, Anaconda-Jupiter notebook. Architecture used in proposed system are given below
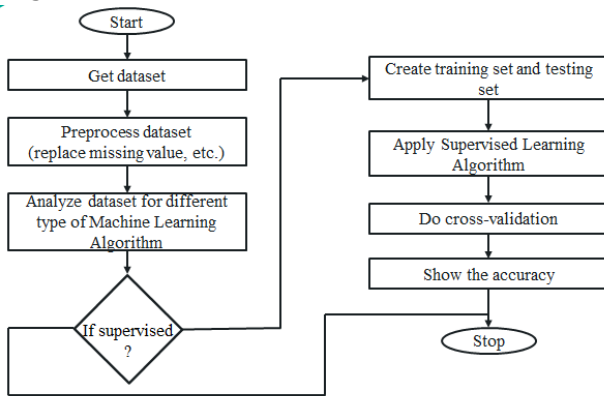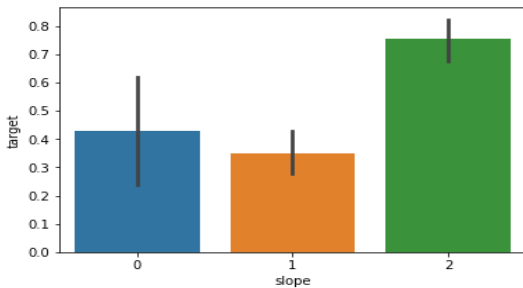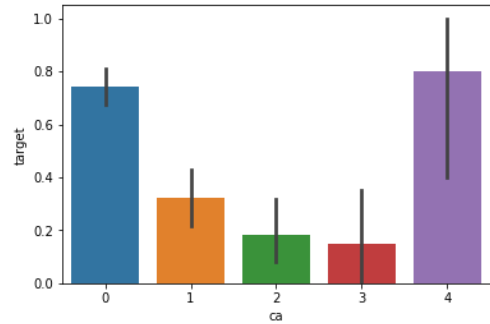

Figure: Architecture of System Design

RESULT AND DISCUSSION

We observe that, stage 2 causes heart pain much more than slope 0 and 1. Number of major vessals colored by flourosopy.

We analyse that people will restcg 1 and 0 are much more likely to have heart disease than restcg 2

Several standard performance metrics such as accuracy, Precision, and error in classification have been considered for the computation of performance efficiency of this model.
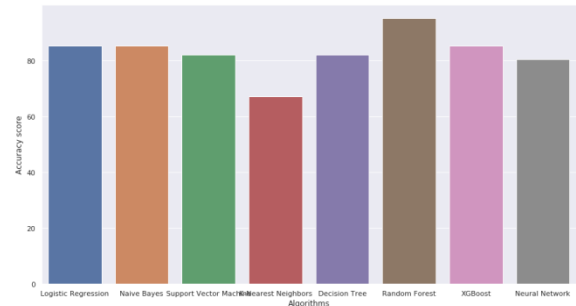

Figure: Performance comparison on various models

CONCLUSION

In this paper, we proposed a method for heart disease prediction using machine learning techniques; these results showed a great accuracy standard for producing a better estimation result. By introducing the new proposed Random forest classification, we find the problem of prediction rate without equipment and propose an approach to estimate the heart rate and condition. Sample results of heart rate are to be taken at different stages of the same subjects; we find the information from the above input via ML Techniques.

BIBLIOGRAPHY

[1] "Prediction of Heart Disease using Artificial Neural Network," VFAST Trans. Softw. Eng., pp. 102–112, 2019

[2] P. Selvakumar and S. P. Rajagopalan, "A survey on neural network models for heart disease prediction," J. Theor. Appl. Inf. Technol., vol. 67, no. 2, pp. 485–497, 2014

[3] A. Methaila, P. Kansal, H. Arya, and P. Kumar, "Early Heart Disease Prediction Using Data Mining Techniques," vol. 6956, no. October, pp. 53–59, 2014

[4] I. S. F. Dessai, "Intelligent Heart Disease Prediction System Using Probabilistic Neural Network," no. 5, pp. 38–44, 2013.

[5] T. Karayilan and Ö. Kiliç, "Prediction of Heart disease using neural network," in 2nd International Conference

on Computer Science and Engineering, UBMK 2017, 2017, pp. 719–723.

[6]M. Mardiyono, R. Suryanita, and A. Adnan, "Intelligent Monitoring System on Prediction of Building Damage Index using Neural-Network," TELKOMNIKA (Telecommunication Comput. Electron Control, vol. 10, no. 1, p. 155, 2015

[7] N. Guru, A. Dahiya, and N. Rajpal, "Decision support system for heart disease diagnosis using Neural Network," 2007.

[8]J. S. Singh Navdeep, "No Title," Intell. Hear. Dis. Predict. Syst. using CANFIS Genet. Algorithm, Int. J. Biol. Med. Sci., no. International Journal of Advance Research, Ideas and Innovations in Technology© 2018, www.IJARIIT.comAll Rights, p. Reserved Page | 987, 2008

[9] R. G. S. Rajkumar Asha, "No Title," Diagnosis Hear. Dis. Using Datamining Algorithm, vol. Issue 10 V, no. Global Journal of Computer Science and Technology, p. Page 38, 2010

[10] Samruddhi Kaware and Dr. V. S. Wadne,JSPM Imperial college of Engineering and Research, Wagholi Pune, "EnhanceCancer and Diabetes Detection by using Machine Learning Techniques" IJRAR March 2020, volume 7, Issue 1.