EasyChair Preprint

№ 9953

Early Prediction of Cardio Vascular Diseases (CVD) Using Artificial Intelligence Technices

Sharad Chand Bind, Pawan Kumar Pradhan and E Rajesh

April 10, 2023

EARLY PREDICTION OF CARDIO VASCULAR DISEASES (CVD) USING ARTIFICIAL INTELLIGENCE TECHNICES

Sharad Chand Bind

Sharadchand350@gmail.com

School of Computer Science & Engineering

Galgotias University Greater Noida, India

Pawan Kumar Pradhan

Pradhanpawan12@gmail.com

School of Computer Science & Engineering

Galgotias University Greater Noida, India

Abstract- Unique from recent times, anticipation of heart disease is one of the most controversial approaches in the medical field. In the advanced stage, one person bites the dust each time due to heart disease. Information science plays an urgent role in handling large amounts of knowledge in the medical field. Since the prognosis for heart disease is a confusing activity, there is a need to change the expected cycle to stay away from related gambling and to warn the patient early. This paper uses the cardiovascular database that is accessible to the UCI AI vault. The proposed work predicts the risk of Heart Disease and demonstrates the patient's gambling rate through a variety of information mining procedures, for example, the Naive Bayes, Decision Tree, Deforestation and Informal Forest. Along with these lines, this paper presents a close-up report by examining the exhibition of various AI statistics. Preliminary results confirm that the Forest calculation achieved a remarkable accuracy of 90.16% compared to other ML calculations made.

Keywords-Decision Tree, Naive Bayes, Logistic Regression, Random Forest, Heart Disease Prediction

INTRODUCTION:-

The proposed work at this paper center is actually about the various mining tests used to anticipate heart disease. The human heart is a vital part of the human body. Basically, it controls the flow of blood throughout our body. Any heart attack can cause problems in various parts of the body. Any kind of interference to the normal functioning of the heart can be called a heart infection. In today's world, heart disease is one of the major causes of death. Heart disease can be due to unhealthy habits, smoking, alcoholism and high fat intake which can cause high blood pressure. According to the World Health Organization, more than 10 million people have been kicked out of the world due to heart disease each year. A strong lifestyle and priority are just ways to prevent heart-related diseases.A key test in current medical care is the best quality management system with the most accurate and effective conclusion. Regardless of whether heartdisease has been identified as one of the leading causes of death on the planet recently, it is also the only one that can be controlled and controlled literally. All accuracy in infection control is in the appropriate time zone for that infection. The proposed work makes an effort to isolate these heart diseases at the beginning of the phase to avoid the dire consequences..

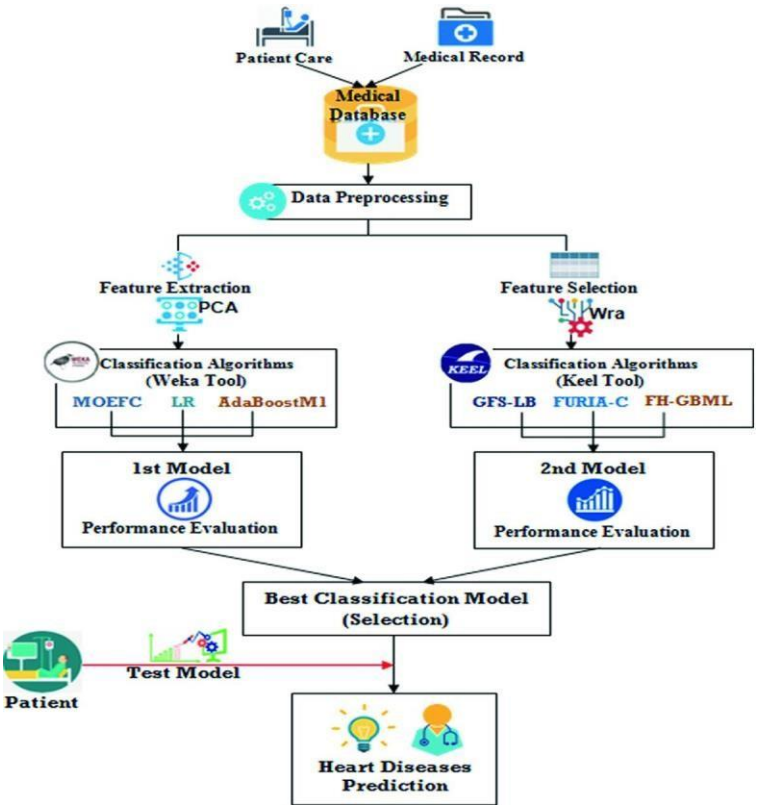
Records of huge arrangement of clinical information made by medical professionals are accessible by undoing and deleting important information from it. Data mining techniques are a way to extract important and hidden data from a large amount of accessible information. For the most part the clinical data set contains different data. Therefore, the flexible use of different information becomes more difficult and more labor intensive. AI (ML) which is a subset of mining knowledge manages a very large amount of data organized in a productive way. In the medical field, AI can be used to diagnose, diagnose and anticipate a wide range of illnesses. The main purpose of this paper is to provide professionals with the tool to identify heartdisease as a first step. This will help to provide effective treatment to patients and avoid serious side effects. ML plays an important role in identifying confidential models and thus separating the information provided. After information research ML techniques help predict heart disease and early determination. This paper provides execution tests of various ML processes, for example, the Naive Bayes, Decision Tree, Deferral Equipment and the Random Forest for expecting heart disease in the first phase.

RELEATED WORK

Calm An important business measure related to the determination of Cardiovascular Heart disease using a machine
Reading statistics have influenced this work. This paper contains a brief summary of writing. The productive prediction of cardiovascular disease is made using various statistics some of which include Logistic Relapse, KNN, Random Forest Classifier etc. It can best be found in the results of the fact that the whole figure has its own unity of registering symbolic goals .The model that integrates the IHDPS has been able to find the limit of choice using the previous model with new AI and in-depth learning. It deals with the most important and basic aspects / information, for example, family relationships related to any heart disease. However, the accuracy found in such an IHDPS model was undoubtedly not really the next new model, for example, diagnosing heart disease using a fake brain organization and differential mechanical calculations and in-depth learning. Gambling genes for coronary heart disease or classified by McPherson et al., Which used a built-in calculation using a few Neural Network procedures and were already ready
wait even if the patient being tested is exposed to a given virus.

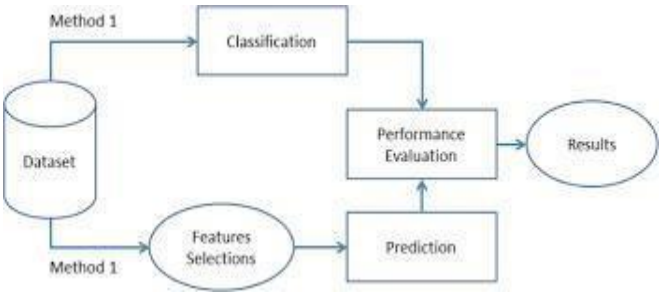
DATA SOURCE

The Fixed Personal Data Set was selected for Keeping as the main concern for their set of heart problems and other illnesses. Coronary disease is a condition in which the heart is affected. According to the World Health Organization (WHO), the highest mortality rate in the elderly is due to heart disease. We take a source of information contained in the clinical history of 304 different patients of different age groups. This database gives us the data we really need for example clinical qualifications, for example, age, resting blood circulation, fasting blood sugar level and other patient information that helps us to differentiate a patient who is willing to have any heart disease or not. This data contains 13 clinical features of 304 patients that help us determine whether or not a patient is at risk for heart disease and assists us in ordering patients at risk for heart disease and those who are not at risk. This Heart Database database is stored in the UCI archive. As shown by this data, an example that encourages the identification of a prone patient is issued for the diagnosis of heart disease. These records are divided into two categories: Preparation and Evaluation. This database contains 303 rows and 14 sections, where each column is compared to a single record



PROPOSED MODEL

The proposed work predicts coronary illness by investigating the previously mentioned four arrangement calculations and does execution examination. The goal of this study is to actually foresee if the patient experiences coronary illness. The wellbeing proficient enters the information values from the patient's wellbeing report. The information is taken care of into model which predicts the likelihood of having coronary illness. Fig. 1 shows the whole interaction included



.Fig. 1: Generic Model Predicting Heart Disease

Logistic Regression:-

Calculated Depletion is the number of orders most of the time used for dual system problems. In reversal of strategy as opposed to inserting a straight line or a high plane, recurring calculations use strategic force to compress the effect of a specific position somewhere in range 0 and 1. There are 13 free features that make counting fun to collect..

Naive Bayes:-

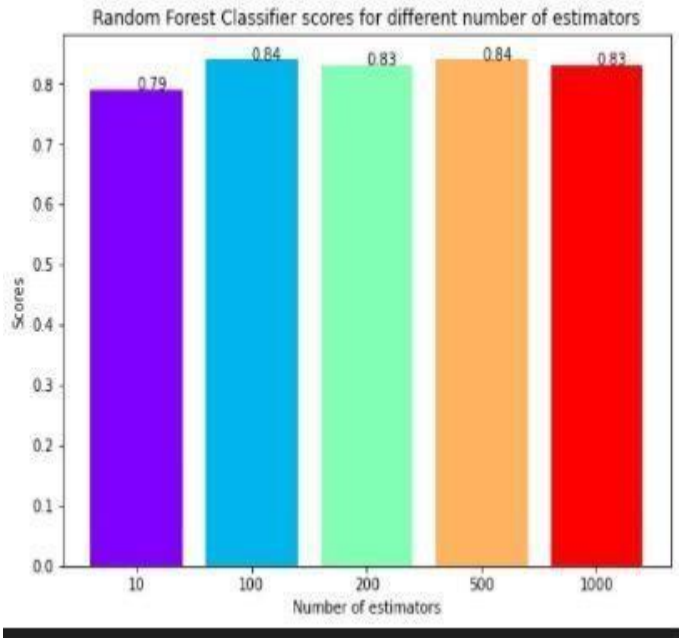
The calculation of Naive Bayes is based on Bayes law []. The freedom between database structures is a fundamental and very important consideration in making estimates. It is easy and quick to expect and very effective when the idea of independence reigns. The Bayes hypothesis confirms the subsequent occurrence of an event (A) if given the previous possibilities of event B targeted by P(A/B)

Classification:-

Random Forest, Decision Tree, Postponement and Naive Bayes order strategies. The information database is divided into 80% of the preparation database and the remaining 20% to the test database. Data setup is a database used to modify a model. The test database is used to view the presentation of the modified model. In all statistics the exhibition is registered and investigated by looking at the various scales used, for example, accuracy, precision, reviews and F scales as shown further. The various statistics investigated in this paper are listed below.

Random Forest

Random Forest algorithms are used for ordering and retrieval. It creates a tree ofknowledge and makes a prediction in the light of that. Extraordinary Forest Calculations can be applied to large databases and can bring the same effect to any event, where record sets of large sets are not available.
Tests produced on the selected tree can be stored for best use in different knowledge. In the jungle there are two stages, the first and most important is tomake unusual backwoods and then do the expectations using the unusual timberland classifier made in the main stage.



RESULT AND ANALYSIS

From these results we can see that although the majority of analysts use a varietyof statistics, for example, SVC, the decision-making tree for patients diagnosed with Heart Disease, KNN, Random Forest Classifier and Logistic relapse produces improved outcomes control. The calculations we used are very accurate, setting aside a ton of money for example generating costs and faster than the calculations used by previous analysts. In addition, the extreme accuracyobtained by KNN and the Random Forest Classifier 84.0% is either too high or equal to the accuracy found in previous tests.

```
print("The score for Random Forest Classifier is {}% with {} estimators.".format(rf_scores[1]*100, [100,
                                                                                                    Python
The score for Random Forest Classifier is 84.0% with [100, 500] estimators.
```

TN True negative: the patient does not have the disease and the test is negative.

FN False negative: the patient has the disease but the test is negative.

In the experiment the pre-processed dataset is used to carry out the experiments and the above mentioned algorithms are explored and applied. The above mentioned performance metrics are obtained using the confusion matrix. Confusion Matrix describes the performance of the model. The confusion matrix obtained by the proposed model for different algorithms is shown below in Table 2. The accuracy score obtained for Random Forest, Decision Tree, Logistic Regression and Naive Bayes classification techniques is shown below in Table 3.

Decision Tree

Decision Tree algorithm is as a flowchart where the inward hub addresses the dataset ascribes and the external branches are the result. Choice Tree is picked on the grounds that they are quick, solid, simple to decipher and very little information arrangement is required. In Decision Tree, the expectation of class mark begins from foundation of the tree. The worth of the root trait is contrasted with records property. On the consequence of examination, the

$$P(A|B) = (P(B|A)P(A)) / P(B) \quad (1)$$

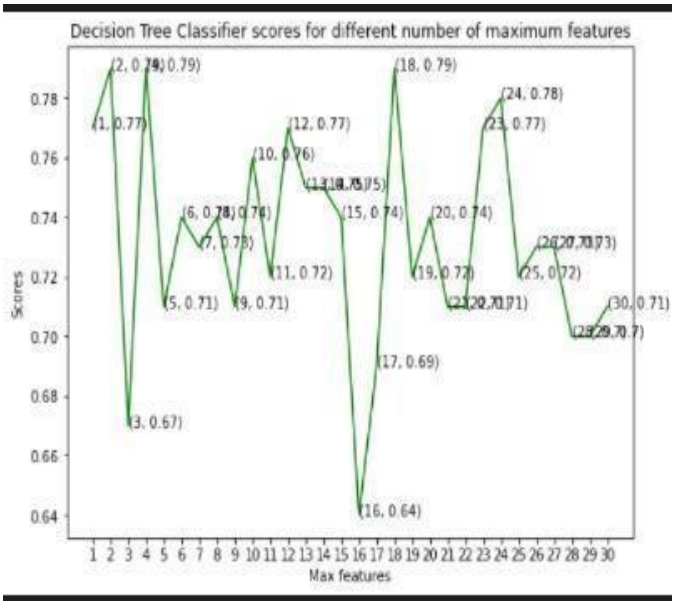


TABLE II. VALUES OBTAINED FOR CONFUSION MATRIX USING DIFFERENT ALGORITHM

Algorithm	True Positive	False Positive	False Negative	True Negative
Logistic Regression	22	5	4	30
Naive Bayes	21	6	3	31
Random Forest	22	5	6	28
Decision Tree	25	2	4	30

TABLE III. ANALYSIS OF MACHINE LEARNING ALGORITHM

Algorithm	Precision	Recall	F-measure	Accuracy
Decision Tree	0.845	0.823	0.835	81.97%
Logistic Regression	0.857	0.882	0.869	85.25%
Random Forest	0.937	0.882	0.909	90.16%
Naive Bayes	0.837	0.911	0.873	85.25%

CONCLUSION

With the growing number of deaths due to heart disease, it is imperative that an effective and accurate cardiovascular prediction system be developed. The aim of the study was to find the most effective ML algorithm for diagnosing heart disease. This study compares the accuracy scores of Decision Tree, Postponement, Informal Forest and Naive Bayes heart predictor algorithms using a UCI machine learning database. The results of this study indicate that the Random Forest algorithm is the most effective algorithm with 90.16% accuracy in predicting heart disease. In the future the work can be improved by creating a web-based Random Forest algorithm and using a larger database compared to those used in this analysis that will help provide better results and assist health professionals in predicting heart disease. effectively and efficiently.

REFERENCES

1. Avinash Golande, Pavan Kumar T, Heart Disease Prediction Using Effective Machine Learning Techniques, International Journal of Recent Technology and Engineering, Vol 8, pp.944-950,2019.

2. T.Nagamani, S.Logeswari, B.Gomathy, Heart Disease Prediction using Data Mining with Mapreduce Algorithm, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-3, January 2019.

3. Fahd Saleh Alotaibi, Implementation of Machine Learning Model to Predict Heart Failure Disease, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 6, 2019.

4. Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin, Design And Implementation Heart Disease Prediction Using Naives Bayesian, International Conference on Trends in Electronics and Information(ICOEI 2019).

5. Theresa Princy R,J. Thomas,Human heart Disease Prediction System using Data Mining Techniques, International Conference on Circuit Power and Computing Technologies,Bangalore,2016.

6. Nagaraj M Lutimath,Chethan C,Basavaraj S Pol.,Prediction Of Heart Disease using Machine Learning, International journal Of Recent Technology and Engineering,8,(2S10), pp 474-477, 2019.

7. UCI, Heart Disease Data Set.[Online]. Available (Accessed on May 1 2020): https://www.kaggle.com/ronitf/heart-disease-uci.

8. Sayali Ambekar, Rashmi Phalnikar,Disease Risk Prediction by Using Convolutional Neural Network,2018 Fourth International Conference on Computing Communication Control and Automation.

9. C. B. Rjeily, G. Badr, E. Hassani, A. H., and E. Andres, Medical Data Mining for Heart Diseases and the Future of Sequential Mining in Medical Field, in Machine Learning Paradigms, 2019, pp. 7199.

10. Jafar Alzubi, Anand Nayyar, Akshi Kumar. "Machine Learning from Theory to Algorithms: An Overview", Journal of Physics: Conference Series, 2018

11. Fajr Ibrahem Alarsan., and Mamoon Younes Analysis and classification of heart diseases using heartbeat features and machine learning algorithms,Journal Of Big Data,2019;6:81.

12. Internet source [Online].Available (Accessed on May 1 2020): http://acadpubl.eu/ap