How to Cite:
Sonkar, S., & Kulkarni, P. (2022). Machine learning approach for high risk prediction of cardiovascular diseases. *International Journal of Health Sciences*, *6*(S2), 10624–10631. https://doi.org/10.53730/ijhs.v6nS2.7828

Machine learning approach for high risk prediction of cardiovascular diseases

Dr. Shrinivas Sonkar
Assistant Professor, Department of Computer Engineering, AVCOR Sangamner
Corresponding author email: sonkar83@gmail.com

Miss. Prajkta Kulkarni
PG Student, Department of Computer Engineering, AVCOR Sangamner
Email: prajaktakulkarni843@gmail.com

***Abstract*---**The most difficult challenge in the field of health sciences is the study of cardiovascular disease. In order to properly diagnose a patient's heart-related issues using this system, it will first be necessary to establish an extensive choice-supporting network. In this article, we presented an efficient system that would be based on the medical histories of patients as well as their previous medical data. The medical industry has made significant strides in recent years in effectively treating people suffering from a variety of infections. For the purpose of achieving an ideal and workable treatment that is also physically supportive, enormous networks can be established to reach an appropriate conclusion. A large number of urgent care centres manage their medical and healthcare services as well as their patients' data and information with the use of medical data frameworks. These therapeutic systems generate enormous amounts of data in the form of photos, text, outlines, and figures. Sadly, this knowledge is being exploited to aid in the expansion of the medical industry. This information contains a substantial amount of hidden data that is not searched, which gives rise to a big question of how to extract useful data from the information or search for hidden data in the information. Therefore, there is an urgent want for the development of an unbelievable medical support that will provide medical professionals with assistance in providing effective therapy for cardiac problems before they occur. The primary purpose of this study is to develop a model that can decide and will provide us information linked with heart problems and vascular disorders based on a previous record of information pertaining to the heart and vasculature information base. It is able to map questions for diagnosing diseases related to the heart or vasculature and assist clinical professionals in reaching appropriate clinical decisions.

***Keywords*---**machine learning, disease prediction, prediction model, classification algorithms, cardiovascular disease.

Introduction

The term "respiratory track failure" is commonly used to refer to a smaller quantity of myocardial infection, and it is the most widely recognized and searched for cause of death in rural, urban, and country settings. The correct diagnosis of respiratory and cardiac failure is essential or primary because unmet expectations can lead to major damage or the heart muscle becoming injured, which is known as cardiomyopathy or cardiovascular disease syndrome, which can result in weakness and death.

When mobile phones become one of the most extensively utilized technological breakthroughs of today, developing an application for supporting coronary illness will give productive and effective results in identifying someone who suffers from chest discomfort. This will lead to the early prediction of cardiovascular and vascular disorders as well as heart attacks, which will lead to the opinion of a physician as well as early diagnosis and therapy as well as early treatment. Even though symptoms like angina pectoris and chest pain are the most typical and visible signs of a heart attack, there are a few more conditions that can easily bring on a heart attack.

In this day and age, the use of modern medicine or the application of modern machine learning algorithms, along with many other modern technologies and pieces of biological equipment, has significantly decreased the overall mortality rate while simultaneously increasing the proportion of healthy people. However, cardiovascular disease (CVD), cancer, chronic respiratory sickness, and diabetes are all extremely lethal conditions, and as a result, they need to be treated appropriately.

Cardiomyopathy and Angina Pectoris, both of which may lead to death, can be prevented by early detection of heart attacks or pulmonary illness. Because of this, early detection is vital. It is critical to be able to foretell a heart attack or a lung ailment. If the blood supply to the coronary artery or the pulmonary artery is abruptly and totally cut off, an acute myocardial infection might occur. Because of this, blood does not reach the heart or the heart zone (also called "heart attack"). A accumulation of fatty deposits, cholesterol, and other byproducts of cellular metabolism may be to blame for this. Some of the heart's muscles might be permanently damaged or even die as a result of blood flow interruptions. If this injury is not addressed early and efficiently when it first emerges, it may be irreparable.

The clinical department has a plethora of information stored, but the primary challenge with medical data mining is the amount of data and complexity of the data, as well as the lack of mathematical categorization and lack of uniform form. In order to extract useful information from the compiled medical data sets, we made use of sophisticated data mining methods. The length of time that passes between the onset of cardiovascular disease or a heart attack and

the patient's decision to seek therapy is a significant challenge. Persons who are oblivious to the signs of a heart attack because they are preoccupied with their daily duties at home or at the workplace, as well as people who live in remote areas, may neglect chest discomfort or chest pain. They may not have the specific purpose to ignore it, but they might squander time deciding whether to see a doctor or end up in the hospital after a certain amount of time has passed. However, timing is of the utmost significance in cases of cardiomyography and heart attacks. Users have access to a wide variety of mobile health tools for the purpose of preventing cardiovascular diseases (CVD), such as mobile apps for self-monitoring or applications connected to healthcare. The state of the art in scientific research has shown that the use of a large number of mobile devices, such as tablets and mobile phones, for communication and feedback, as well as smartphone apps, is beneficial. Due to the fact that the medical diagnosis of heart disease is very important, despite the fact that the task is difficult and expensive, we proposed a medical diagnosis system in order to improve the quality of medical treatment provided to patients suffering from heart and vascular conditions while simultaneously reducing costs. People will be able to determine whether or not they are at risk for heart failure either at that moment in time or at a later point in time if we are successful in providing a service that is both widespread and viable and sustainable.

Related Work

M. A. Jabbar et.al, [1] alternative way of employing affiliation management processes in the clinical sector has been presented in order to identify heart disease predictions or detect vascular disease. A significant quantity of information on medical services is collected by the sector. Unfortunately, these services have not been used in order to identify encased data that is successful. If you use abstract examples, you will see that hidden examples and connection softening have not been employed to their maximum potential. The issue may be solved with the use of information mining methods. Data mining has shown a variety of uses in a variety of fields, including science and commerce. Information mining focuses a lot of attention on affiliation rules, arrangements, and aggregation. These are all very significant topics.

Ms. M. C. S. Geetha et.al, [2] studied the widely used classification methods included in the medical data set, which contribute to the prediction of cardiovascular illnesses, which are the leading causes of mortality throughout the globe. To accurately forecast heart attacks based on experience and expertise, which might be challenging, doctors need to work with trained specialists. The modern healthcare industry is full with confidential yet important information that is used to make choices. The results of the trials that were conducted demonstrate this method. J48, SIMPLE CART, and REPTREE have, as was to be predicted, a better level of predictive accuracy and precision than other algorithms.

M. Akhil Jabbar et.al, [3], it was pointed out that the approach of design confirmation known as the nearest neighbour (KNN) is one that is simple, well-

known, competent, and powerful. KNN is a kind of direct classifier in which the arrangement of pieces is determined by the category to which they belong based on their immediate surroundings. The nature of the clinical information base is one that is expansive. It's possible that grouping will yield fewer accurate results if the data set has an excessive number of irrelevant qualities. In India, coronary heart disease has the highest mortality rate of any illness. The leading cause of mortality in Andhra Pradesh is coronary heart disease, which accounts for 32% of all fatalities. This percentage is comparable to that of both Canada (35%) and the United States. Consequently, it is important to define the kind of emotional support network that will best suit your needs. The physician will make significant advancements as a result of this option. Another technique for achieving a strong order is developed and fixed in this study. This strategy combines KNN with genetic programming. Genetic techniques execute global inquiries in complicated large and multi-mode contexts and produce appropriate arrangements.

Chaitrali S. Dangare [4] presented the anticipated results and framework of heart illness based on the improving quality of the information. In order to determine the chance of a patient having heart disease, this study makes use of clinical criteria such as gender, strain on the circulatory system, cholesterol levels, blood pressure, and credit history. Up until very recently, 13 characteristics were employed for making predictions. The task of inspection also incorporates two more duties, including toughness and smoking, among others. The evaluation of the data mining system was performed on the data pertaining to pulmonary or coronary sickness. This included the use of specialized opinion and decision trees, naive Bayes, Random Forest, and neural networks.

Zeinab Arabasadi et.al, [5] a hybrid diagnostic approach with increased accuracy and precision was suggested as a means of identifying coronary artery disease and pulmonary disease. In point of fact, the enhancement of the neural network's initial weights with the help of genetic algorithm enables the suggested technique to bring about an improvement of roughly 13 in terms of both the accuracy and performance of the neural network.

Sahar H. El-Khafifand Mohamed A. El-Brawany [6] emphasised the fact that the ECG signal is well-known for its non-linear shifting nature and serves as the primary instrument for carrying out this examination. The non-linear component of its constituents is subject to greater natural change under normal and abnormal situations combined, compared to conditions that are linear. This study employs the one-dimensional shear from awful regions, which is more challenging for usual subjects as a result of the high availability, measurement (HOS) that is necessary to retain the phase information. The feed forward multi-layer neural system (NN) and machine learning has a Negligent Backlash (BP) learning approach, and it is utilised as a computerised ECG classifier to determine the likelihood of diagnosing coronary heart disease based on typical ECG signals.

Senthil Kumar Mohan and colleagues [7] projected that heart disease would turn out to be the worst illness on the planet. On a yearly basis, heart disease (cardiovascular arrhythmia) is responsible for the deaths of 9.4 million individuals, 62 percent of which are the result of strokes and 37 percent of which are the result of coronary heart disease. The subsequent assessment of different cardiac illnesses at an early stage is a very crucial thing to do in order to determine the severity of the condition. Finding cardiovascular arrhythmias requires using current conventional ECG examination methodologies (such as RR span, wavelet mutation, and group computation) (for example, support vector machines, K closest neighbours, and Levenberg Marquardt neural networks). Using these applications will not result in an accurate identification of the issue, even if a huge number of relevant details have been extracted.

I. S. Siva Rao, T. Srinivasa Rao [8] heart disease was anticipated to be the most harmful illness in the globe. On a yearly basis, heart disease (cardiovascular arrhythmia) is responsible for the deaths of 9.4 million individuals. Of these deaths, 62 percent are the result of strokes, and 37 percent are the result of coronary heart disease. Consequently, determining the severity of the different cardiac problems at an early stage is a very crucial thing to do. Finding cardiovascular arrhythmias requires the use of current conventional ECG examination methodologies (such as RR span, wavelet mutation, and group computation) (for example, support vector machines, K closest neighbours, and Liebenberg Marquardt neural networks). Even after collecting a vast number of different pieces of information, the issue detection that these tools provide will not be correct.

Amma, N.G.B [9], It is hypothesized that medical diagnostic systems carry out significant work in clinical settings and are used by medical professionals for the purposes of diagnosis and therapy. The clinical evaluation system is represented in this study by the terms normal or high risk of cardiovascular infection. The functionality of this structure is achieved by combining specific genetic programming with the preferences of the sensory system as a whole. The multi-layer feed forward kind of neural system has had its settings specifically modified for the complex layout issues that have been presented. Genetic approaches may be used to tackle the load problem of the nervous system since these methods can locate an appropriate load arrangement on medical diagnostic in a shorter amount of time.

Saba Bashir et.al, [10], this article discusses the use of data science to the prediction of cardiac illnesses in the realm of health care. In spite of the significant number of studies that have been undertaken on this topic, there is still room for advancement in terms of the accuracy of predictions. As a result, the primary emphasis of this research is on feature selection and collaborative approach methods and algorithms. In this study, numerous data sets on heart disease are utilized for experimental analysis, and the results demonstrate that utilizing algorithms results in greater accuracy.

Proposed System

An efficient system for predicting cardiovascular attacks will be proposed by us. This mechanism will, first, learn deep characteristics, and then, using machine learning techniques, will train these learnt features. The findings indicate that the classifier is superior to any and all other classifiers when all characteristics and the same training sample are used for the training process. This not only demonstrates that the performance gain is statistically significant, but it also shows that the accuracy is greater than the system that was used before. It is particularly difficult to forecast heart disease or attacks using low population, high-dimensional data sets since the samples are not sufficient to acquire the accurate mapping between attributes and category labels. In the majority of cases, this work is completed by manually developing and choosing characteristics in the existing literature. It has been discovered that random forest, in comparison to other technologies, is able to determine the fundamental structure of the data.

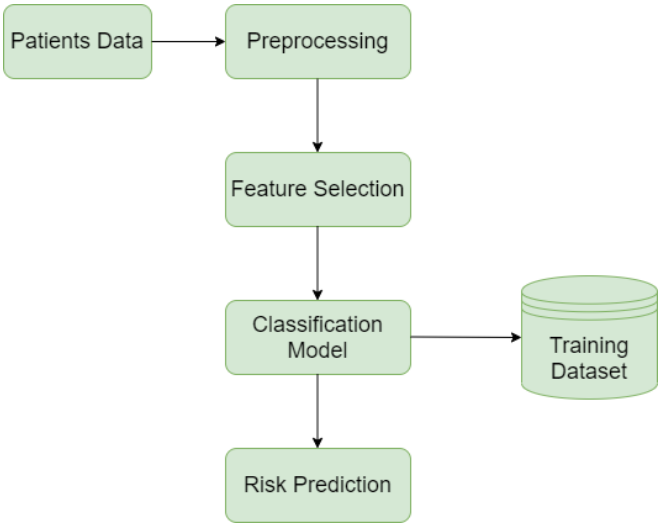


Figure 1: System Architecture

Algorithm:
Classification Algorithm:
“Random Forest:
Input:
Step 1: Upload dataset
Step 2: Symptoms is the set of input attributes
Step 3: cardiovascular disease prediction is the set of output attributes
Step 4: sample is a set of training data
Function Iterative Dichotomies returns a decision tree

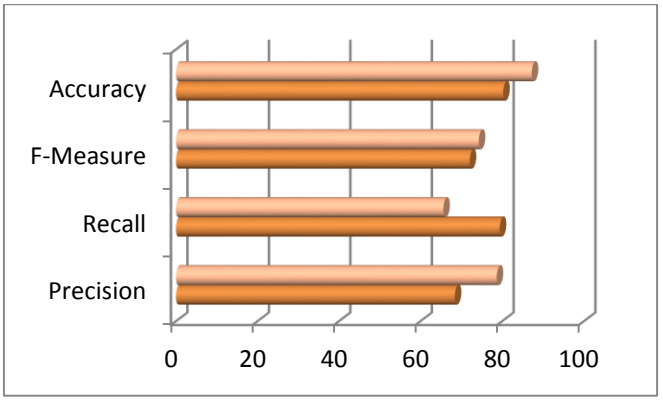
1. Create root node for the tree
2. If (all inputs are positive, return leaf node positive)
   If Else (if all inputs are negative, return leaf node negative)
   Else (Some inputs are positive, and some inputs are negative,
   Check condition

(Positive>negative||Positive<negative), then return result)
3. Calculate the entropy of current state H(S)
4. For each attribute, calculate the entropy with respect to the attribute 'X' denoted by H(S, X)
5. Select the attribute which has maximum value of IG(S, X)
6. Remove the attribute that offers highest value from the set of attributes
7. Repeat until we run out of all attributes or the decision tree has all leaf nodes.

Output:
Dataset value will be retrieved".

Result and Discussion

We analyzed the accuracy of the suggested system for predicting cardiovascular illness based on the number of training samples, and we displayed the comparison's findings visually. The graph and table that follows illustrate the accuracy of the decision tree classification approach applied to the prediction of cardiovascular disease.



	Naïve Bayes	Random Forest
Precision	68.45	77.70
Recall	79.44	65.64
F-Measure	72.11	74.31
Accuracy	80.29	88.26

Conclusion

As part of this body of work, we have presented a technique for the categorizing of coronary disease as well as pulmonary disease. We have a plan to have the planned technique approved; it involves including the patient's cardiac testing result. Train informative collections that are taken from the datasets that are available online. Both Naive Bayes and Random Forest are considered to be aggressive classification approaches within the realm of machine learning. Our methodology makes use of both of them. With the aid of this prediction model,

medical professionals may undertake an accurate and effective diagnostic procedure for cardiovascular disease using less patient characteristics. At the present time, cardiac sickness is acknowledged as the primary contributor to death across India and in the state of Andhra Pradesh. Distinguishing the evidence of important risk factors is one way that we may bring down the overall mortality rate by using our cardiovascular disease system.

References

1. M. A. Jabbar, B.L Deekshatulu, Priti Chandra, “An evolutionary algorithm for heart disease prediction”CCIS,PP 378-389 , Springer(2012).
2. Ms.M.C.S.Geetha, Dr.I.Elizabeth Shanthi, Ms.N. Sanfia Sehnaz,” Ana-lyzing the Suitability Of Relevant Classification Techniques On Medical Data Set For Better Prediction” International conference on I-SMAC 2017.
3. M.Akhil jabbarB.L Deekshatulua Priti Chandra International“Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm”Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013.
4. Chaitrali S Dangare “Improved Study Of Heart Disease Prediction System Using Data Mining Classification Techniques”, International Journal Of Computer Applications, Vol.47, No.10 (June 2012).
5. Zeinab Arabasadi , Roohallah Alizadehsani , Mohamad Roshanzamir , Hossein Moosaei , Ali Asghar Yarifard “ Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm” Science Direct 2017.
6. Sahar H. El-Khafifand Mohamed A. El-Brawany, “Artificial Neural Network-Based Automated ECG Signal Classifier”, 29 May 2013.
7. Senthilkumar Mohan, Chandrasegar Thirumalai, And Gautam Srivastava” Effective Heart Disease Prediction Us-ing Hybrid Machine Learning Techniques”IEEE ACCESS 2019
8. I. S. Siva Rao, T. Srinivasa Rao, “Performance Identification of Different Heart Diseases Based On Neural Network Classification”.ISSN 0973-4562 Volume 11, Number 6 (2016) pp 3859-3864.
9. Amma, N.G.B “Cardiovascular Disease Prediction System using Genetic Algorithm”, IEEE International Conference on Computing, Communication and Applications, 2012.
10. Saba Bashir, Zain Sikander Khan, Farhan Hassan Khan, Aitzaz Anjum, Khurram Bashir,” Improving Heart Disease Prediction Using Feature Selection Approaches”, International Bhurban Conference on Applied Sciences Technology 2019.