The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 267 - 275

A Clinical Support System for Prediction of Heart Diseases using Machine Learning Techniques

Prof. Smita S. Wagh[1], Amruta Javir[2], Gayatri Zambare[3], Vaishnavi Patil[4], Neha Potdar[5]

[1] Prof. Smita S. Wagh, Dept. Of Computer Engineering, Pune, Maharashtra, India[2]Amruta Javir, Dept. Of Computer Engineering, Pune, Maharashtra, India, [3]Gayatri Zambare, Dept. Of Computer Engineering, Pune, Maharashtra, India [4]Vaishnavi Patil, Dept. Of Computer Engineering, Pune, Maharashtra, India [5]Neha Potdar, Dept. Of Computer Engineering, Pune, Maharashtra, India

Abstract -- In recent times, heart disease prediction is one of the complicated and major cause of death throughout the world. In this era, approximately one person dies per minute due to heart disease. This cannot be easily predicted by the medical practitioners as it is a difficult task which demands expertise and higher knowledge for prediction. An automated system in medical diagnosis would enhance medical efficiency and also reduce costs. We will design a system that can efficiently discover the rules to predict the risk level of patients based on the given parameters about their health. The goal is to extract hidden patterns by applying data mining techniques such as Convolutional neural network (CNN), Support Vector machine(SVM), Decision Tree and Random forest which are noteworthy to heart diseases and to predict the presence of heart disease in patients where the presence is valued on a scale. Predicting heart disease requires massive amounts of data that are too complex to process and analyse by conventional methods. Our goal is to find machine learning techniques that are computationally efficient and accurate in predicting heart disease. Data mining combines statistical analysis machine learning and database techniques to extract hidden patterns and relationships from large databases.

Keywords -- Convolutional Neural Network (CNN), Decision Tree, Naive Bayes, Random Forest, Machine Learning, Heart Disease Prediction

I. Introduction

According to the World Health Organization, 12 million people worldwide die each year from cardiovascular disease. Heart disease is one of the leading causes of morbidity and mortality in the world population. Cardiovascular disease prognosis is one of the most important issues in data analysis. Over the past few years, the worldwide burden of cardiovascular disease has increased dramatically. Several studies have been conducted to pinpoint the factors most influencing cardiovascular disease and accurately predict overall risk. Heart Disease is even highlighted as a silent killer which causes the death of a person without obvious symptoms. The early diagnosis of heart disease plays a vital role in making decisions on lifestyle changes in high-risk patients and in turn, reduces complications.

Machine learning proves to be effective in assisting in making decisions and predictions from the large quantity of data produced by the healthcare industry. This project aims to predict

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 267 - 275

future Heart Disease by analyzing patients' data that classifies whether they have heart disease or not using the machine-learning algorithm.

Machine Learning techniques can be a boon in this regard. Even though heart disease can occur in different forms, there is a common set of core risk factors that influence whether someone will ultimately be at risk or not. By collecting the data from various sources, classifying them under suitable headings finally analysing them to extract the desired data we can say that this technique can be very well adapted to the prediction of heart disease.

The work proposed in this paper focuses mainly on various data mining practices employed in heart disease prediction. The human heart is the principal part of the human body. It regulates blood flow throughout our body abnormalities in the heart can cause pain in other parts of the body. When the normal functioning of the heart is disrupted, it can be classified as heart disease- Cardiovascular disease can be caused by an unhealthy lifestyle, smoking, drinking, and increased fat intake, which can lead to arterial hypertension. According to the World Health Organization (WHO), more than 10 million people die yearly from cardiovascular disease worldwide. A healthy lifestyle and early detection are the only ways to prevent cardiovascular disease.

A major challenge in modern healthcare is providing the highest quality services and effective and accurate diagnoses. Heart disease can also be effectively controlled and managed, despite becoming the leading cause of death in the world in recent years. All precision in disease management lies in the timely detection of this disease. Analyse massive medical data records generated by healthcare professionals and extract valuable insights. Data mining technology is a means of extracting valuable and hidden information from large amounts of available data. Essentially, medical databases consist of individual pieces of information.

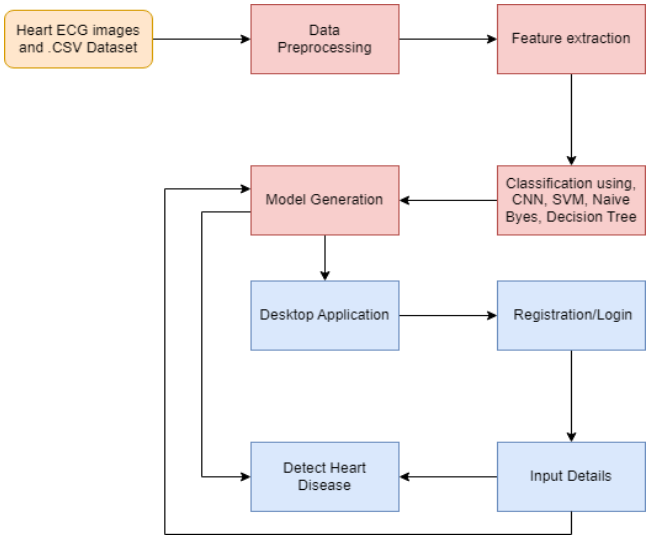
Decision-making with discrete data thus becomes a daunting task.

This paper presents performance analysis of various ML techniques such as Naive Bayes, Decision Tree, Convolutional Neural Network, Support vector Machine and Random Forest for predicting heart disease at an early stage.

II. Proposed Method

It is a desktop-based machine learning application that is trained with the help of UCI dataset. The functioning of the system begins with data collection and the selection of important attributes. Then the required data is pre-processed into the required format. Input data is then further divided into two parts i.e. training data and testing data.

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 267 - 275



Fig(I). Architecture Diagram

A. Methodology

The accuracy of the system is obtained by testing the system using test data. The implementation of thesystem occurs as follows:

1) Dataset - Provide dataset (This means that the data collected should be made uniform and understandable for a machine that doesn't see data the same way as humans do.

2) Pre-processing - Real-world data generally contains noises, and missing values, and may be in an unusable format that cannot be directly used for machine learning models. Data pre-processing is required task for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model.

3) Feature Extraction - Feature Extraction aims to reduce the number of features in a dataset by creating new features from the existing ones (and then discarding the original features). These new reduced set of
features should then be able to summarize most of information contained in the original set of features.

4) Classification - The Classification algorithm is a Supervised Learning technique that is used to identify the category of new observations on the basis of training data. In Classification, a program learns from the
given dataset or observations and then classifies new observation into a number of classes or groups.

The algorithm will calculate the probability of presenceof heart disease.

The complete description of the 14 attributes used in the proposed work is mentioned in Table(I) shown below.

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 267 - 275

Id. No.	Attribute Description	Distinct Values of Attribute
1.	Age(age) - represents age of a person in days	Multiple integer values (29 to 77)
2.	Sex(sex) – it represents the gender of person (0-Female, 1-Male)	0,1
3.	Chest Pain (cp) – Represents the severity of chest pain of thepatient	0,1,2,3
4.	Cholestrol (chol) — It represents the cholesterol level of the patient	0,1,2,3
5.	FBS (fbs) – It represents the patient's fasting blood pressure	0,1
6.	Resting ECG (restecg) – It shows the result of ECG	0,1,2
7.	Thalach (thalach) – It represents maximum heart rate of the patient	71-202
8.	Exang (exang) – It represents the exercise induced angina	0: No 1: Yes
9.	Oldpeak (oldpeak) – It represents depression level of the patient	0-6.2
10.	Slope (slope) – It represents patient's condition during peakexercise	0,1,2
11.	CA(ca) – Flouroscopy result	0,1,2,3,4
12.	Thallium stress test (thal) – Testrequired for patient who is suffering from chest pain and breathing problem	0,1,2,3
13.	Target – It represents the final result from the dataset (0 - less possibility of heart disease, 1- high chances of heart disease)	0,1

Table(I)-Features selected from the dataset

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 267 - 275

B. Algorithms

1) Convolutional Neural Network

Convolutional Neural Networks specialized for applications in image and video recognition. CNN ismainly used in image analysis tasks like Image recognition,

Object detection Segmentation. There are Four types of layers in Convolutional Neural Networks:

1) Convolutional Layer: In a typical neural network each input neuron is connected to the next hidden layer. In CNN, only a small region of the input layer neurons connect to the neuron hidden layer.

2) Pooling Layer: The pooling layer is used to reduce the dimensionality of the feature map. There will

be multiple activation pooling layers inside the hiddenlayer of the CNN.

3) Flatten: Flattening is converting the data into a

1-dimensional array for inputting it to the next layer.

We flatten the output of the convolutional layers to createa single long feature vector.

4) Fully-Connected layer: Fully Connected Layers form the last few layers in the network. The input to the fully connected layer is the output from the final Pooling or Convolutional Layer, which is flattened andthen fed into the fully connected layer.

2) Support Vector Machine

In machine learning, support-vector machines (SVMs, also support-vector networks) are supervised learning models with associated learning algorithms that analyse data used for classification and regression

analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making

it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on the side of the gap on which they fall.

3) Decision Tree

Decision is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. Decision Tree is chosen because they are fast, reliable, and easy to interpret, and very little data preparation is required in a Decision tree,

there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.

4) Naive Bayes

Naive Bayes algorithm is a supervised learning algorithm, which is based on the Bayes theorem and used for solving classification problems. It is mainly used in text classification

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 267 - 275

that includes a high-dimensional training dataset. We are using this algorithm, as it helps in building fast machine learning models that can help in making the predictions quicker.

III. Literature Survey

[1] Aditi Gavhane et al proposed a paper "Prediction of Heart Disease Using Machine Learning", in which training and testing of dataset is performed by using neural network algorithm multi-layer perceptron. In this algorithm there will be one input layer and one output layer and one or more layers are hidden layers between these two input and output layers. Through hidden layers each input node is connected to output layer. This connection is assigned with some random weights. The other input is called bias which is assigned with weight based on requirement the connection between the nodes can be feedforwarded or feedback.

[2] Aakash Chauhan et al, proposed "Heart Disease Prediction using Evolutionary Rule Learning". Data is directly retrieved from electronic records that reduce the manual tasks. The amount of services are decreased and shown major number of rules helps within the best prediction of heart disease. Frequent pattern growth association mining is performed on patient's dataset to generate strong association.

[3] Anjan N. Repaka et al, proposed a model stated the performance of prediction for two classification models, which is analyzed and compared to previous work.The experimental results show that accuracy is improved in finding the percentage of risk prediction of our proposed method in comparison with other models.

[4] Lakshmana Rao et al, proposed "Machine Learning Techniques for Heart Disease Prediction" in which the contributing elements for heart disease are more. So, it is difficult to distinguish heart disease.To find the seriousness of the heart disease among people different neural systems and data mining techniques are used.

[5] A. Lakshmanarao, Y. Swathi, P. Sri Sai Sundareswar proposed a paper "Machine Learning Techniques for Heart disease prediction" this paper contains machine learning methods for heart disease detection. As raw datasets contain unbalanced samples of class distribution, we applied three sampling techniques on the dataset. After applying sampling techniques accuracy and recall rates increased drastically. For random oversampling, SVM given the best accuracy. For Synthetic Minority Oversampling, Random Forest and Extra tree Classifier given the best accuracy. For Adaptive synthetic sampling, Random Forest and Extratree Classifier given the best accuracy.

[6] C. Sowmiya, Dr.P. Sumitra proposed a paper "Analytical Study of Heart Disease Diagnosis Using Classification Techniques" this paper contains analysis several classification techniques that are very useful in data mining for detecting the heart disease. From the above existing system, we have analysed all types of heart disease diagnosis. We have classified several techniques that are useful in data mining for detecting heart disease. It has proven that classification based techniques contribute high effectiveness and obtain high accuracy compare than the previous methods. Future work it can experiment for another algorithm for a priori algorithm.

[7] Aadar Pandita, Siddharth Vashisht, Aryan Tyagi, Prof. Sarita Yadav proposed a paper "Review Paper on Prediction of Heart Disease using Machine Learning Algorithms" this

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 267 - 275

paper contains majority of researchers have used the Cleveland Heart Disease Dataset available from the UCI repository containing 76 attributes and 303 instances, of whichonly 14 attributes are used due to missing values.

There are huge benefits to having feature selection methods so as to minimize the number of attributes that one has to use in order to build an accurate model by checking the correlation between various attributes and their impact on the accuracy of the models. It can be seen from various research papers in the field that KNN and Neural Network works quite accurately in most cases for the prediction of heart diseases.

[8] Raunak Verma,, Shashank Tandon, Mr. Vinayak proposed a paper "Heart Disease Prediction using Machine Learning". In this paper, the purpose was to describe the various ML techniques that are useful in predicting heart disease. Effective and accurate predictions with small number of characteristics and evaluation of the purpose of this study. The data was previously processed and used in the model. K-NearestNeighbor, Random Forest and Support Vector Classifier algorithms working very well. We can continue to expand this research that integrates other ML strategies such as time series, integration rules and integration with other integration strategies. Considering the limitations of this study, there is a need to use complexity and combination of models to achieve high accuracy in predicting early heart disease.

[9] Gaurav Meena,Pradeep Singh Chauhan,Ravi Raj Choudhary proposed a paper "Empirical Study on Classification of Heart Disease Dataset-its Prediction and Mining.This paper contains application of data mining techniques in HDD is an emerging trend in the world. It has attracted the attention of medical practitioners and academics. This paper has identified ten articles related to application of data mining techniques in HDD, and published between 2006 and 2016. It endeavors to provide a research review on the application of data mining in the HDD domain and methods which are most often used. Even though this review cannot claim to be thorough, it does give reasonable imminent and shows the frequency of research on this subject.

[10] Dr. Manoj S. and Yuvaraju B. proposed a paper titled "Design and Implementation of Cardiac Disease Prediction Using Machine Learning." Predictive results indicate user status and lead to diagnosis. With recent technological advances, machine learning algorithm have evolved significantly, and the proposed system uses multiple algorithms for efficiency and accuracy. Also, the algorithm provides a near and reliable output based on the user's input. As more people use thissystem, their current heart condition will become more recognized, and the rate of death from heart disease willeventually decrease

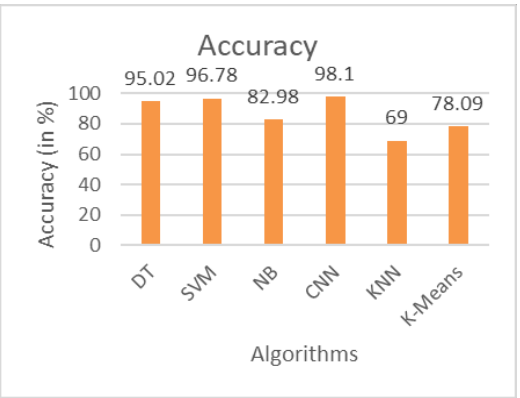
IV. Results And Evaluation

A heart disease prediction system checks for diseases that can be associated with a patient. Provides results based on patient's input. Various experiments are performed and the above algorithms are studied to compute the performance and validation of the developed models using the preprocessed datasets to run the experiments. Results are computed using confusion matrices and their accuracies are compared using different algorithms. We have performed several tests with different percentages. The graph from Fig(II) shows the results obtained comparing the accuracy of the SVM, DT, NB, CNN, KNN, and K-Means algorithms. From

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 267 - 275

the Table(II) below, it is clear that the classification accuracy of SVM, DT, NB, and CNN is higher than other A KNNs and K-Means. However, the accuracy of CNN is the best compared to all compared algorithms. Evaluation depends on the data used in the training and test sets.

The table shows that for the proposed model, CNN performed best with 98.1% accuracy, and SVM came second with 96.78% accuracy. To obtain the highest accuracy in predicting heart disease, we used CNN for image preprocessing of the proposed model. We provide both text and image datasets to the proposed model as inputs for predicting probable heart disease. As a result, it predicts heart disease in patients.



Fig(II). Comparison of the algorithms

ML Algorithms	Accuracy(in %)
DT	95.02
SVM	96.78
NB	82.98
CNN	98.1
KNN	69
K-Means	78.09

Table(II)-Accuracies of the algorithms

V. Conclusion

First, the algorithms were implemented. The dataset was trained separately for all algorithms. They were all then tested. The most efficient algorithm should be selected based on various criteria. We found the CNN algorithm to IEEE

be the most efficient with an accuracy of 98.1%. The accuracy of decision tree and SVM was 95.02% and 96.78% respectively, and the accuracy of NB was 82.98%. Therefore, these four algorithms are further implemented with a better user interface. We used Python's Spyder IDE to create an interactive desktop application. Equipped with machine learning algorithms, this desktop application then builds robust models for predicting heart disease. This helps the

end user to make a preliminary prediction of the heart condition. As heart disease is one of the leading causes of death in India and around the world, applying promising technologies like machine learning for early prediction of heart disease will have a significant impact on society by usimg this model.

VI. References

[1] Aditi Gavhan , Gouthami Kokkula, Isha Pandyaand Prof. Kailas Devadkar ; “Prediction of HeartDisease Using Machine Learning”,2018.

[2] C.Sowmiya , Dr.P.Sumitra ;. “: Analytical Study of Heart Disease Diagonosis Using ClassificationTechniques.,”2017

[3] Aakash Chauhan,Aditya Jain ,Purushottam Sharma,Vikas Deep; “Heart Disease Prediction using Evolutionary Rule Learning.”,2018

[4] Gaurav Meena ,Pradeep Singh Chauhan, Ravi Raj Choudhary; “Empirical Study on Classification ofHeart Disease Dataset Prediction and Mining.”,2017

[5] Dhabliya, D., & Sharma, R. (2019). Cloud computing based mobile devices for distributed computing. International Journal of Control and Automation, 12(6 Special Issue), 1-4. doi:10.33832/ijca.2019.12.6.01

[6] Lakshmanarao,Y.Swathi, P.Sri SaiSundareswar; “Machine Learning Techniques ForHeart Disease Prediction”,2019

[7] Anjan Nikhil Repaka, Sai Deepak Ravikanti , Ramya G Franklin ; “Design And Implementating Heart Disease Prediction Using Naives Bayesian.”,2019

[8] Norma Latif fitriyani, Muhhammad syafrudin

[9] , Gunjar Alfian and Jongtae rhee; “An Effective Heart Disease Prediction Model for a clinical Decision Support System,2020

[10] Aadar Pandita1, Siddharth Vashisht, Aryan Tyagi, Prof. Sarita Yadav; “Review Paper on Prediction of Heart Disease using MachineLearning Algorithms”,2021

[11] Manoj S1, Dr. Yuvaraju B.N; “Design & Implementation of Heart Disease Prediction using Machine Learning”,2020

[12] Dhabliya, D., Soundararajan, R., Selvarasu, P., Balasubramaniam, M. S., Rajawat, A. S., Goyal, S. B., . . . Suciu, G. (2022). Energy-efficient network protocols and resilient data transmission schemes for wireless sensor Networks—An experimental survey. Energies, 15(23) doi:10.3390/en15238883

[13] Raunak Verma, Shashank Tandon ,Mr. Vinayak;“Heart Disease prediction using Machine learning”,2016

[14] Tülay Karaylan, ozkan kille ; “Prediction of HeartDisease using Neural network”,2017

[15] Priyanka S. Sangle1, R. M. Goudar2, A.N. Bhute ; “Methodologies and Techniques for Heart Disease Classification and Prediction”,2020