International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 05 | May 2023 www.irjet.net p-ISSN: 2395-0072

Prediction of Heart Disease Using Machine Learning and Deep Learning Techniques.

Spurthi S R[1], Shree Subeeksha G[2], Keerthi M S[3], Kavya P[4], Dr. Gururaj Murtugudde[5]

[1, 2, 3, 4]*Student, Dept. of Computer Science and Engineering, REVA University, Bangalore, India*
[5]*Professor, Dept. of Computer Science and Engineering, REVA University, Bangalore, India*

Abstract - *The primary cause of death has historically been heart related disease worldwide over past few decades, thus it is crucial and worrisome to anticipate any such disorders. Heart-related disease diagnosis and prognosis is a difficult task that calls for greater accuracy, correctness, and perfection since a small error can result in weariness or even death, which has a significant global impact. Due to a multitude of risk factors, such as smoking, diabetes, high cholesterol, and similar conditions, it can be challenging to diagnose heart disease. As a result of these circumstances, it is urgent to develop precise, practical, and trustworthy methods for making an early diagnosis, as doing so will benefit people everywhere by enabling them to receive the necessary therapy before the condition worsens. The data from the dataset is obtained using contemporary methods like data mining and machine learning techniques, and the fetched data is then utilised to forecast cardiac disease. With the help of deep learning techniques like CNN and MLP as well as machine learning methods like ADABOOST and EXTRATREES, this work attempts to predict the likelihood of getting cardiac illnesses.*

***Key Words*: Machine Learning, Adaboost, ExtraTree, Deep Learning, CNN, MLP, Heart disease Prediction.**

1. INTRODUCTION

The most important crucial organs in the human body, the heart plays a crucial part in the blood's circulation throughout the body. Heart disease can be caused by a number of things, such as unhealthy lifestyle choices, drinking, smoking, and job-related stress. This syndrome may result in abnormal heart blood flow and led to severe conditions such as strokes, coronary heart disease, and heart attacks. If the heart is faulty in any manner, cardiac disorders such congenital heart disease, heart failure, and arrhythmia may also manifest. Heart disease, is major cause for illness and loss of lives globally, killing 12 million people each year according to the World Health Organisation. Predicting cardiovascular disease is so essential, and many researchers have investigated the most important risk factors to precisely determine overall risk. The prevention of heart-related illnesses depends on heart illness being detected early.

The enormous data created by medical industry is used by machine learning algorithms to make predictions and judgements. One such application is the analysis of patient information to identify patients with heart disease, forecast future heart illness, and identify it early. Different types of heart problems can be diagnosed, detected, and predicted with the aid of machine learning methods like Adaboost, ExtraTree algorithms, and deep learning techniques like CNN and Multi-layer Perceptron (MLP). The basic risk factors for heart disease are universal across the range of heart illnesses, enabling patients to get appropriate care and avert negative outcomes. To identify hidden patterns and analyse data to identify heart sickness at an early stage and prevent consequences, machine learning is essential.

2. LITERATURE REVIEW

In this study, Kuldeep Vayadande et al. [1] used the 303-row and 14-attribute UCI heart dataset and implemented ml algorithms such as logistic regression, XGBoost, and random forest, which had good accuracy ratings of 88.52% in comparison to all other models. The accuracy of deep learning algorithms like MLP and ANN is 86.89% and 85.25 percent, respectively.

In a study to predict cardiovascular illness, Shafique R et al. used a variety of classification methods, such as Extra Tree Classifier, Logistic Regression, SVM, and NB The Cleveland heart dataset, which has two classes and 13 attributes, was used. The research discovered that, out of all the machine learning methods examined, Extra Tree Classifier achieved the highest accuracy rate of 90%.

A. Lakshmanarao et al. predicted the risk of developing a coronary heart disease for a period of ten years in patients using Framingham Heart Study dataset. To solve the issue of an unbalanced dataset, three distinct sampling techniques were used using the dataset, which had 15 features. Using random oversampling, the study discovered that SVM was the most accurate machine learning model, while using adaptive synthetic sampling and synthetic minority sampling, ExtraTree and Random Forest were determined to be the most accurate models. This study reveals how ml techniques forecast a person's risk of getting heart disease, which may help with the early detection and treatment of this common condition.

Shadab Hussain et al. recommended a 1D convolutional neural network (CNN) architecture to detect cardiac disease. The Cleveland dataset was used to train and test the model,

which has 13 features. Training accuracy for the model was 97%, and test accuracy was 96%. The 1D CNN architecture surpassed all other classification algorithms when the performance of each was examined in the study, including few ML algorithms and ANN. This study shows the potential of 1D CNNs for cardiac disease prediction, which could help with early diagnosis and treatment of this condition.

Sayali Ambekar and Rashmi Phalnikar[5] to extend the work of heart disease prediction, they have used CNN based on unimodel illness risk prediction algorithm on the heart dataset of UCI repository with 12 attributes. The CNN-UDRP algorithm have achieved an accuracy of more than 65%. They have considered 500 iteration and input layer which contains 10 input factors in CNN algorithm to obtain an accurate result. They aim to show that CNN-UDRP algorithm performance on structured data for disease risk prediction.

3. MATERIALS AND METHODS

3.1 Description of Data

The Cleveland, Statalog, Hungarian, Swiss, and Long Beach VA heart datasets were pooled in this study to produce a larger dataset with a total of 1190 occurrences. 11 common features and a target variable were included in the merged dataset. Six nominal and five numerical variables were included in the dataset, which was used to analyse and forecast heart disease. The accuracy of machine learning models used to forecast cardiac disease may be enhanced by the use of larger datasets with diverse samples.

Table-1 : Dataset Attribute Description

Observation	Type	Description
Age	Numeric	Patient Age
Sex	Nominal	Gender of patient (M-1, F-0)
Chest_Pain	Nominal	Type of chest pain
Resting_BP	Numeric	Blood Pressure at rest
Cholesterol	Numeric	Serum cholesterol
Fasting_Blood_Sugar	Nominal	Fasting blood sugar > 120 1 - true 0 - false
Resting_ECG	Nominal	Electrocardiogram at rest.
Max_Heart_Rate	Numeric	Maximum recorded rate of heart
Exercise_Angina	Nominal	Induced Angina by excercise
Old_peak	Numeric	ST depression induced by exercise relative to rest
ST_slope	Nominal	ST measured slope of peak exercise
Target	Numeric	1- Heart diagnosed with disease 0 - Patient is normal

3.2 Algorithms Used

Four machine learning algorithms were utilised in this study to predict cardiac disease. Deep learning algorithms, a branch of machine learning, are used in this situation. Adaboost Classifier and ExtraTrees Classifier are the two machine learning algorithms utilised, and Convolutional Neural Network (CNN) and Multilevel Perceptron Classifier (MLP) are the two deep learning methods employed.

i. Adaboost:

A machine learning algorithm for categorization tasks is called the AB. To create a more reliable overall classifier and combines number of weak learners, or simple models. The first weak learner in the method is trained using the data before the error is calculated. Adaboost also offers a unique approach of machine learning: as ensemble learning tool, it builds on the core notion that many effective learners can outperform a single effective learner.

ii. Extra-Trees Classifier:

Extremely Randomised Trees Classifier, also referred to as Extra-Trees Classifier, is a type of ensemble learning method which combines results of various de-correlated decision trees. Each of these decision trees seeks to differentiate samples from different classes in the target by removing impurities in some way. It simply differs conceptually from a Random Forest in terms of how forest's Decision Trees are constructed.

iii. Convolutional Neural Network(CNN):

CNN is a artificial neural networks used in deep learning to analyse visual data. Given that CNN has an input layer, an output layer, numerous hidden layers, and millions of parameters, it can learn complicated objects and patterns. CNN uses a stacking technique for convolutional layers.

iv. Multi-layer Perceptron(MLP) :

There are a minimum of three layers in a neural network, an input layer, a hidden layer, and an output layer. Each input and each output have a single neuron, also called as a node. The number of hidden levels and nodes in each hidden layer is up to you. In this system, the input layer receives input signals from the outside world and sends them to every neuron in the hidden layer.

4. PROPOSED WORK

A person can die from heart disease without exhibiting any overt symptoms, which is why it is sometimes referred to as a silent killer. The nature of the sickness is the cause of growing concern about the condition and its severe implications. So efforts to predict the future development of this terrible disease in the past still exist today. As a result, numerous techniques and technologies are often evaluated to suit the demands of contemporary health.

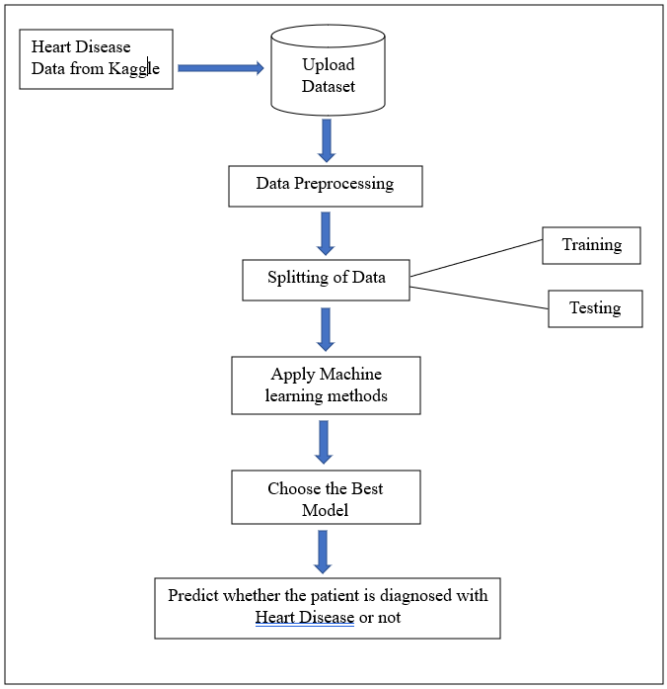

Fig 1. Workflow Diagram

In this study, we used a heart dataset from Kaggle that contains 1190 instances with 1 target column and 11 common variables including age, sex, type, heart rate, cholesterol, etc. Later on in the data preprocessing approach, duplicate and missing values are checked. The dataset had no duplicate or missing values. The dataset was then divided into 20% for testing and 80% for training. The dataset is exposed to different machine learning and deep learning algorithms applications, including Adaboost, Extra-Tree, CNN, and MLP, where CNN achieves highest accuracy of any technique. Multiple layers, including Sequential, Conv2D, MaxPool2D, Flatten, Dropout, and Dense, is utilised to train the CNN model. 'Softmax' is used as an activation function with an output layer to stack a couple more layers. This model is used to forecast heart disease since it had an accuracy rate of 98.28%.

5. RESULT AND DISCUSSION

Both machine learning and deep learning algorithms have demonstrated impressive performance in this paper, according to our study. The heart dataset, which contained information on a total of 1190 instances with 12 attributes, was used in the implementation. Adaboost, a machine learning method, obtained an accuracy of 89.08%; ExtraTrees, an accuracy of 94.12%. This model's performance improved with the addition of deep learning techniques like MLP and CNN. Multi-layer Perceptron accuracy is 83.61%, and Convolutional Neural Network (CNN) accuracy with 98.28%. CNN was chosen as the most effective algorithm for heart disease prediction as a result.

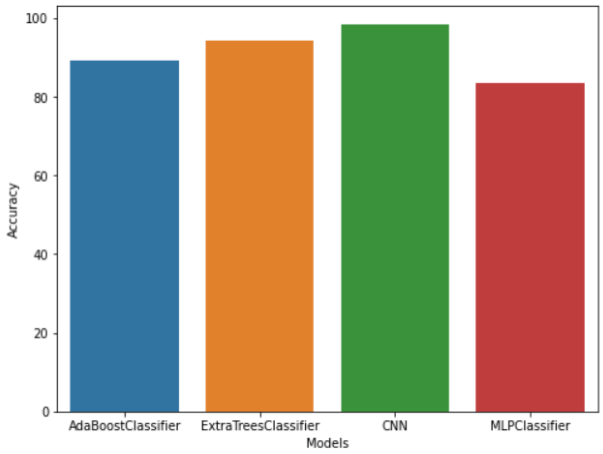

Fig 2. Accuracy Comparison

6. CONCLUSION AND FUTURE WORK

The rise in fatal heart disease cases, it is imperative to develop a system that can accurately and successfully predict heart diseases. To enhance the prediction of cardiac illness, this study applies a number of machine and deep learning techniques. The current study demonstrated classification using a sizable sample of participants. Our research leads us to the conclusion that the CNN algorithm performs the best for combined multiple cardiac datasets among all other methods. The suggested method is implemented as a computer software system, which makes it easier to comprehend and gain a better understanding of the person's cardiac health as soon as feasible. The future study should focus on predicting heart disease using lesser number of clinical parameters, so that everyone may readily learn about their heart health and take immediate precautions. And still more deep learning algorithms to be explored to know about the better performance of the models. The study should involve the usage of more heart dataset as of now this study includes US, UK, and other European cities dataset where it should be concentrated on Asian heart dataset too, so that the system would help for maximum population.

REFERENCES

[1] Kuldeep Vayadande, Rohan Golawar, Sarwesh Khairmar, Arnav Dhiwar, Sarthak Wakchoure, Sumit Bhoite, Darpan Khadke, "Heart Disease Prediction using Machine Learning and Deep Learning Algorithms", 1st

International Conference on Computational Intelligence and Sustainable Engineering Solution (CISES 2022), doi:10.1109/CISES54857.2022.9844406

[2] Rahman Shafique, Arif Mehmood, Saleem ullah, Gyu Sang Choi, "Cardiovascular Disease Prediction System Using Extra Trees Classifier", 2019, doi: https://doi.org/10.21203/rs.2.14454/v1

[3] A. Lakshmanarao, Y.Swathi, P.Sri Sai Sundareswar, "Machine Learning Techniques For Heart Disease Prediction", International Journal of Scientific & Technology Research ,Volume 8, Issue 11, 2019

[4] Shadab Hussain, Santosh Kumar Nanda, Susmith Barigidad, Shadab Akhtar, Md Suaib, Niranjan K.Ray, "Novel Deep Learning Architecture for Predicting Heart Disease using CNN", 2021 19th OITS International Conference on Information Technology, doi: 10.1109/OCIT53463.2021.00076

[5] Sayali Ambedkar, Rashmi Phalnikar, "Disease Risk Prediction by Using Convolutional Neural Network", 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 978-1-5386-5257-2/18

[6] Chavana Sateesh, R. Balamanigandan, " Heart Disease Prediction using Innovative Decision tree Technique for increasing the Accuracy compared with Convolutional Neural Networks", 2022 2nd International Conference on Innovative Practices in Technology and Management(ICIPTM), doi:10.1109/ICIPTM54933.2022.9754196

[7] Muhammad Haseeb Aslam, Syed Fawad Hussain, "Prediction of Heart Failure by using Machine Learning and Feature Selection", 2022 International Conference on Emerging Technologies (ICET) 2022, doi:10.1109/ICET56601.2022.10004668

[8] Dr. C.N.Vanitha, Dr. K. Vanitha, C.Narmatha, S.A.Krishna, R.Dhivakar, " Heart Disease Prediction using Enhanced Deep learning",2022 International Conference on Applied Artificial Intelligence and Computing(ICAAIC), doi:10.1109/ICAAIC53929.2022.9793175

[9] CH Raja Shaker, Anisetti Sidhartha, Anto Praveena, A. Chrsity, B.Bharati, "An Analysis of Heart Disease Prediction using Machine Learning and Deep Learning Techniques", 2022, Sixth International Conference on Trends in Electronics and Informatics(ICOEI), doi:10.1109/ICOEI53556.2022.9776745

[10] Fadoua Khennou, Charif Fahim, Habiba Chaoui, and Nour El Houda Chaoui, "A Machine Learning Approach: Using Predictive Analytics to Identify and Analyze High Risks Patients with Heart Disease",2019 International Journal of Machine Learning and Computing, Vol.9, No.6, December 2019, doi:10.18178/ijmlc.2019.9.6.870