OPEN ACCESS

Volume: 11

Special Issue: 1

Month: July

Year: 2023

E-ISSN: 2582-0397

P-ISSN: 2321-788X

Impact Factor: 3.025

Received: 01.05.2023

Accepted: 15.06.2023

Published: 01.07.2023

Citation:
Subburaj, T., and GR Gagan. "Heart Disease Detection with Data Analytics and Machine Learning." *Shanlax International Journal of Arts, Science and Humanities*, vol. 11, no. S1, 2023, pp. 95–101.

DOI:
https://doi.org/10.34293/sijash.v11iS1-July.6322

Heart Disease Detection with Data Analytics and Machine Learning

Dr. T. Subburaj
Department of Master of Computer Applications
Rajarajeswari College of Engineering

Gagan G.R
Department of Master of Computer Applications
Rajarajeswari College of Engineering

Abstract

ML(Machine Learning) has recently played an important part in the healthcare business, and among all It is important diseases, heart among the more typical crucial and vital diseases to forecast. The quantity of instances is rapidly increasing every day. Considering that it has been discovered four persons older than or equal to 30 and 50 have a stroke every minute, We are using methods of machine learning to alleviate this problem. The heartdisease dataset was utilised by Kaggle for this project. This research examines and shows the likelihood of developing utilising several machine learning classifying methods, like Naive Bayes, to identify cardiovascular diseases Random Forest, SVM, and others. Later on, the Stacking Ensemble Learning Technique is utilised in order boost efficiency of our classification models.

Keywords: Cardiology, Data Analysis, Data Mining, Diseases, Pattern Classification, Support Vector Machine, Heart Disease Prediction, ArtificalNueral Network (ANN).

Introduction

Cardiovascular among the more typical common diseases that might shorten a person's lifetime nowadays. Every year, 17.5 million individuals die as a result of heart disease. Because in the heart a fundamental aspect of our bodies, life is contingent on its component functioning. heart condition is a condition that impairs The goal is the heart. Many parts of health promotion and clinical practise rely on an assessment of a person's risk over cardiovascular conditions disease. A risk prediction model can be developed from a longitudinal research using multivariate regression analysis. Because digital technologies are continually evolving, healthcare facilities keep massive substantial data in their databases, which are extremely complicated and hard to analyse. Data harvesting methods, and moreover in the investigation of, automated learning techniques are crucial. Many types of data in medical facilities.

The approaches and algorithms may be applied directly to a dataset to create models or to derive important conclusions and inferences from the dataset. Age, Sex, and Frequently existing danger signs for heart disease. Fasting blood pressure, kind of chest pain, resting ECG (test that monitors electrical movement in the heart), number

of main vessels coloured by fluoroscopy Fasting blood sugar, Threst Blood Pressure (high blood pressure), Serum Cholestrol (determine a possibility of evolving heart disease), Thalach (maximum heart rate achieved), ST depression (finding on an electrocardiogram, trace when the ST segment is abnormally low below the baseline), painloc (chest pain location (substernal=1, otherwise=0)), Exang (exercise with angina), smoke Hypertension, Obesity, obesity, and eating habits.

Cardiovascular disorders, particularly illness inner heart (CVDs), is the main contributor to global morbidity and mortality, which accounts for more than 70% of all fatalities. Conferring to The impact of illnesses worldwide Study, CVD causes almost 43% of all mortality. 2017. Unhealthy diet, smoke, excessive sugar, and being overweight or obese are all frequent danger signs a cardiac condition in high-income nations. However, the prevalence of chronic illnesses is increasing in low- and middle-income nations. Between 2010 and 2015, the global economic burden of CVDs was expected to be roughly USD 3.7 trillion.

Literature Survey

In medical centres Different data mining techniques and machine learning approaches have used in numerous attempts to develop sickness prediction systems. algorithms. K. Polaraju et al.[7] suggested Prediction a Heart Condition Using a Multiple linear regression is demonstrated by the multiple regression model. is effective at predicting the risk of developing heart disease. The work is done with a database for training of 3000 examples with the previously specified 13 distinct properties. Using The data set is divided into two sections, with 30% for testingand 70% of the informationused for training parts on the basis of data, There is no doubt about the Regression. method outperforms the other algorithms pertaining to classification accuracy.

Marjia et al.[8] created a forecasting tool for cardiovascular disease utilising KStartogether with j48, SMO, Bayes Net, and multilayer perception using WEKA software. Using k-fold KStar, Multilayer Perception, and J48 methods are outperformed by cross validation, SMO and Bayees Net in terms of performance from different factors. The precision performances of the Algorithms continue to unsatisfactory. As a hence, the precision of the performance is increased further to provide better decisions for medical diagnosis.

S. Seema and co. [9] focus Using (ANN)Artificial Neural Networks , Naive Bayes, Decision Trees, (SVM)Support Vector Machines , and to predict chronic disease using previous health information (ANN). A comparison research is conducted on classifiers one has a higher accuracy rate, to be able to compare them. SVM displays the most accuracy in this trial.
rate, but for diabetes, Naive Bayes has the highest accuracy.

The authors include Ashok Kumar Dwiveedi and others. [10] sugested many methods KNN, Classification Tree, and Naive Bayes, among othersSVM,Logistic Regression and ANN. When compared Comparatively, Logistic Regression offers higher accuracy.
MeghaShahi et al. sugested the System for Predicting Cardiovascular Disease Employing Data Analysis Methods [11].In healthcare facilities, WEKA Software is utilised for automated sickness diagnosis and to provide service quality. SVM, Several algorithms were incorporated into the study, such as Naive Bayes, KNN,Association rule, ANN, and Decision Tree. work. The report suggested that SVM is more effective and accurate than various data mining techniques.

Existing System

The current approach predicts cardiac utilising a variety of machine learning approaches, such as Naive Bayes, Random Forest, SVM,and others, and compares their accuracy ratings.

The present mechanism for heart disease forecasting employs Naive Bayes, Random Forest, and Stability Vector Machines (SVM). In These approaches are called machine learning often

employed for classification problems. The system analyses patient information such as age, gender, blood pressure, and cholesterol levels and then uses algorithms to forecast the chance of heart disease. Despite the success these algorithms have considerable success, they do have significant limits that may compromise their correctness and usefulness.

The present system produced a heart disease algorithm called as StackingClassifier based on Ensemble Learning Technique, where several classification models will be constructed and trained using their performance on the training set to compare their accuracy scores. Stacking Ensemble Learning will be employed later to enhance the classification model's efficacy.

Overfitting: Sometimes ensemble learning approaches overfit a set of practise data, resulting in excellent accuracy however, poor performance on a set of practise data fresh, unknown data. This can lead to inaccurate forecasts and poor patient outcomes.

Limited Interpretability: Because ensemble learning approaches can be complicated and difficult to comprehend, understanding the underlying elements that contribute to anticipate cardiac illness can be difficult. This may impede physicians' and healthcare professionals' capacity to apply the model in decision-making procedures.

Data Bias: Ensemble methods to learning are susceptible to data bias, which occurs once the model has been trained on unbalanced datasets, resulting in biassed predictions. The outcome of this may affect the accuracy and dependability of the forecasts.

Ensemble learning approaches may be computationally costly, requiring large processing resources to train and optimise. This might limit their usefulness. in resource-constrained areas with limited access to high-performance computer resources.

Need for skill: Because ensemble learning approaches demand skill in data science and machine learning to build and optimise, they are inaccessible to non-experts.

Proposed System

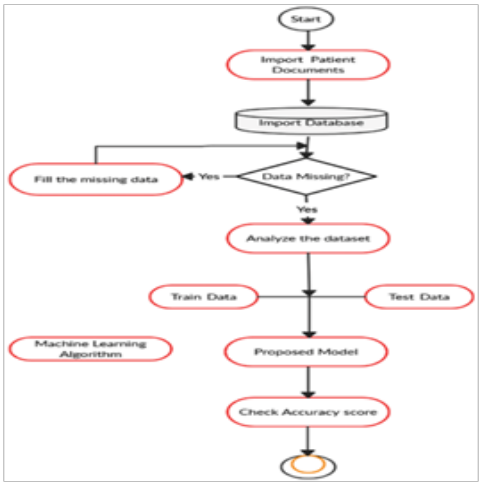

Figure 1 Proposed System

This is the proposed System of this project flow diagram which will show project flow from end to the last. According to the World Health Organisation (WHO), 17.7 million people worldwide die abruptly as a result of cardiovascular diseases each year. The capacity to anticipate the complexity of their health at an early stage may assist people, owing to the heart disease prediction system.

Traditional approaches for predicting heart illness include a doctor's assessment or a battery of medical tests such as an ECG, a stress test, a cardiac MRI, and so on. There is a large amount of hidden information in the existing health care data. Having access to this concealed knowledge aids in making sound judgements. Computer-based data as well as contemporary data mining methods employed to get acceptable results.

Existing methods predict outcomes effectively, but additional data qualities and the complexity of health indicators offer the foundation for the advancement of novel solutions.When predicting cardiac disease in people, we use synthetic neural networks (ANN). this suggested method. The project "Heart Utilising synthetic neural systems to predict disease (ANN)" intends to use ANNs to produce a model for early Cardiovascular disease forecasting has been identified.ANNs are complex machine acquiring algorithms for can discover patterns and correlations in data, making them an excellent candidate for forecasting complicated medical problems such as heart disease. The suggested system is built on the UCI machine learning repository's Cleveland cardiac disease dataset / Kaggle.After that, the data is pre-processed and utilised to train an ANN model using supervised learning techniques. The algorithm has been optimised to forecast a person's risk of developing heart disease patients with great accuracy.The trained model is then assessed on a separate dataset for performance and accuracy. Finally, the project intends to provide a simple to use interface that will doctors and healthcare workers to enter patient data and obtain expected results.The suggested methodology has the capacity to raise the accuracy and speed with which cardiac disease is diagnosed, allowing for earlier intervention and improved patient outcomes. This initiative has the potential for significantly affect public health by to the creation of contributing to the advancement of AI-powered healthcare solutions.

Implementation

Suggested method, we Cardiovascular disease can be predicted with the aid of Machine learning, specifically (ANN)Artificial Neural Networks. The suggested technique entails creating an ANN prediction design capable reliably identify peoplewith cardiovascular disease risk. The approach employed ANNs to estimate the likelihood of heart illness using the prominent dataset Cleveland Heart illness, which is the repository for machine learning at UCI Accessible on Kaggle. Because ANNs can learn from big datasets and handle noisy and complicated data, they are an excellent candidate for medical diagnosis.

Creating an ANN model that is capable to preprocessed data and reliably forecast the likelihood a cardiac condition in patients in the proposed system. An artificial brain network (ANN), often known as a neural network, is a mathematical model based on biological brain networks. The creation of the artificial neural network according to research into the human brain. The human brain is a complex network of neurons. Axons, dendrites, and synapses are components of neurons. The planned ANN is composed on three layers : a source layer, a concealed layer, and a result layer. The suggested system is built with the Flask web framework, and the trained ANN model is deployed as a web application that accessible to anyone user, healthcare professional, or patient to forecast the likelihood of heart disease.

Detection of Heart Disease Early: The ANN-based prediction model may detect heart disease enabling its beginnings early therapies and perhaps saving lives.

Accuracy: ANNs can learn from enormous datasets and find complex patterns and correlations in the data, resulting in more accurate heart disease forecasts.

Speed: Because the ANN-based prediction model can process data fast, it can provide speedy findings Which has applications timely interventions and treatment.

Customizability: The ANN-based predictive model may be tailored to the patient's age, gender, and medical history, resulting in more personalised forecasts and better outcomes.

Scalability: The ANN-based prediction model may be scaled up to analyse enormous datasets, allowing for population-level analysis and identification of heart disease risk factors.

Reduced Expenses: Using the ANN-based prediction model, early diagnosis of heart illness can minimise healthcare expenses connected with additional extreme forms of heart disease.

Improved Patient Outcomes: The ANN-based prediction modelling can enhance patient results and lower death rates by identifying people at cardiovascular risk factor for disease and delivering early therapies.

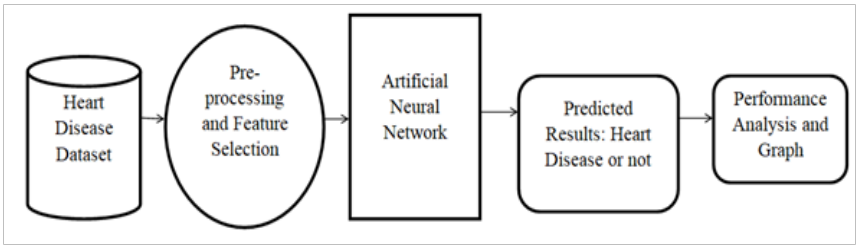

Figure 2 Flow Diagram

This is the flow diagram of this project by seeing this diagram it will easy to understand the work flow of the project.

The suggested system has the maximum accuracy of 91.1%, making it the best among existing system models. OnImageNet, the suggested system model may be pre-trained.

Methodology

Machine Learning

Machine learning is a developing technique that allows computers to learn autonomously from historical data. Machine learning employs a number of algorithms to construct mathematical models and make predictions based on past data or information. It is being utilised for a number of activities including image recognition, audio recognition, email filtering, Facebook auto-tagging, recommender systems, and many more.

This machine learning course introduces you to machine learning and moreover many (ML) machine learning techniques like supervised, unsupervised, and reinforcement learning. Regression and classification models, clustering approaches, hidden Markov models, and other sequential models will be covered.As it actually world, we are enveloped by individuals who have the potential should learn from their mistakes. own experiences and computers or robots that function on our orders. Can a machine, like a person, learn from prior experiences or data? So now comes Machine Learning's part.

Artificial Neural Network (ANN)

The phrase "Artificial Neural Network" comes from biological neural networks, which create the structure of the human brain. Artificial neural networks, like the human brain, include neurons that are coupled to one another in various levels of the networks. These neurons are referred to as nodes. In the science of artificial intelligence, an (ANN)Artificial Neural Network seeks to duplicate the network of neurons that makes up the human brain so that computers can understand things and make decisions in a human-like manner.

Computers are programmed to act just like linked brain cells in the artificial neural network.

Results

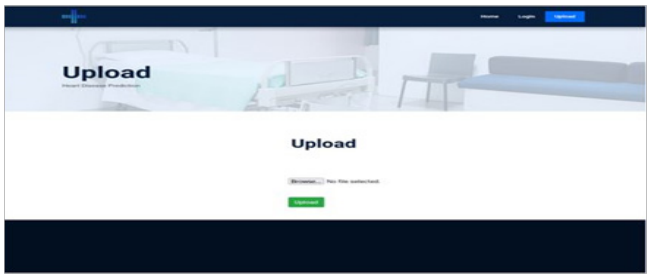

Figure 3 Upload:-The CSV File will be Uploaded

After the user is logged to his or her account then he need to upload the CSV file where the CSV file should contain some data about the patient which may include age, sugar level before and after fasting etc .It will be uploaded to test some cases. Then it will give some result about the data that have been given.

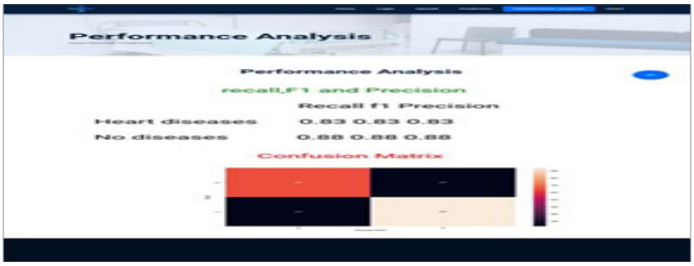

Figure 4 Performance Analysis:-Analysing the Heart Disease

This snap shot is the Performance analysing graph which will be containing the success ratio of showing whether there is a heart disease or no disease ,This will be showed in graph format to make understand easily.

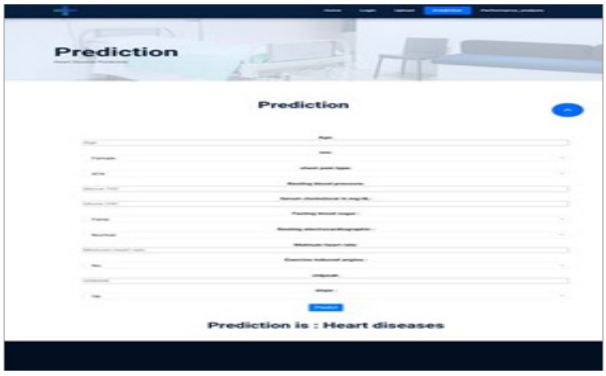

Figure 5 Prediction of Disease by giving Some Values

After uploading the CSV File ,Some data which are present in the CSV file has to be given ,The CSV file will be having so many data up to thousand,If we need extra then we can add it, After giving some values the data will be analysed to check whether there is heart disease in the patient or not, In the above snap shot there is a result of heart disease.

Conclusion

Major influencing elements for determining cardiac illness and numerous research efforts in predicting There are heart conditions. and described in this study. discovered that not all qualities are considered by every researcher. Few characteristics are deleted by a few researchers to capable of deliver more accuracy. We had a thorough discussion on the major issues of various research studies for The prognosis for heart illness is still unknown. solved. In the future, researchers should use an effective algorithm to incorporate they're all elements for identifying heart disease. The analysis of the literature revealed the need for combinational and more complex models to enhance the accuracy of forecasting the early beginning of cardiovascular illnesses.

The study presents a method for accurately predicting cardiac disease by integrating the use of artificial neural networks with support vector machines. Making use of the Cleveland Heart Disease database, we want to give guidance to prepare for the test and the system, resulting in the most efficient model of the numerous rule-based combinations. Furthermore, our research suggests a comparison examination of accuracy. Furthermore, the most effective and well-balanced model could identified

References

1. SharmilaS et al., ""Data Mining Techniques for Analysis of Heart Disease Prediction", International Journal of Advanced Networking & Applications (IJANA), Volume: 08, Issue: 05, Pages: 93-95 (2017).
2. Chaitrali S. Dangareetalbetter research into cardiovascular disease detection system utilising information mining classification Techniques", International Journal of Computer Applications, Volume: 47, Pages: 44-48 (2012).
3. AnimeshHazra, Subrata Kumar Mandal, Amit Gupta, Arkomita Mukherjee and Asmita Mukherjee, "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review", Advances in Computational Sciences (2017) Vol 10, No 7, ISSN 0973-6107, Pages 2137–2159.
4. https://www.techopedia.com/definition/30597/predictive- data-mining
5. http://www.heart.org/HEARTORG/Conditions/HeartAttack/WarningSignsofaHeartAttack/Warning-Signs-of-a- HeartAttack_UCM_002039_Article.jsp#.WNpKgPl97IU.
6. www.who.int/cardiovascular_diseases/en/.
7. http://food.ndtv.com/health/world-heart-day-2015-heart-disease-in-india-is-a- growing-concern- ansari-1224160.
8. Innovative Research in Science, Engineering and Technology, Volume 3, Special Issue 3, March 2014. Sunil Ray, "Learn Naïve Bayes algorithm" and "Decision tree- Simplified", URL: www.analyticsvidhya.com, Retrieved
9. Hlaudi Daniel Masethe et al., "Prediction of Heart Disease using Classification Algorithms", Proceedings of the World Congress on Engineering and Computer Science [WCECS], Volume: II, ISBN: 978-988-37253-7-4 ISSN: 2078-0958 (Print), ISSN: 2078-0966 (Online) (2014).
10. InternationalConference on Computing for Sustainable Global Development (INDIACom), IEEE, Pages: 3099-3103 (2016).