IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 7, July 2023

DOI: 10.17148/IARJSET.2023.107105

Detection of Cardiovascular Disease Applying Machine Learning and Data Analytics

Varshitha C[1], Dr. H K Madhu[2]

Student, MCA, Bangalore Institute Of Technology, Bangalore, India[1]

Associate Professor, MCA, Bangalore Institute of Technology, Bangalore, India[2]

Abstract: Machine Learning has recently played an important part in the healthcare business, and among all of the major diseases, sentiment ailment is one of the furthermost crucial and vital diseases to forecast. The quantity of instances is rapidly increasing every day. Because it has been initiate that four individuals amongst the ages of 30 and 50 have a stroke every minute, we are spread over appliance scholarship procedures to alleviate this problem.

The heart disease dataset was utilised by Kaggle for this project. This research examines and displays employing a variety of machine-learning categorization techniques, such as naive Bay Probabilistic Forest, and SVM, in identifying cardiovascular conditions, and others. Later on, the Stacking Ensemble Learning Technique is utilised to progress the routine of our classification models.

Cardiology, Data Analysis, Data Mining, Diseases, Pattern Classification, Support Vector Machine, Heart Disease Prediction, Artifical Nueral Network (ANN).

I. INTRODUCTION

Vascular disease is one of the furthermost common ailments that might shorten a person's lifetime nowadays. Every year, 17.5 million individuals die as a upshot of heart disease.Because the heart is a fundamental aspect of our bodies, life is contingent on its component functioning. Heart ailment is a condition that impairs the occupation of the heart.Many parts of health promotion and clinical practise rely on an assessment of a person's risk for coronary heart 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 volumes of data in their databases, which are extremely complicated and difficult to analyse.Data mining techniques and machine learning systems are critical in the homework of various 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 Gender are common hazard aspects for heart disease. Fasting blood pressure, kind of chest pain, resting ECG (test that monitors electrical activity in the heart), number of main vessels coloured by fluoroscopy Fasting blood sugar, Threst Blood Pressure (high blood pressure), Serum Cholestrol (determine the risk of unindustrialized heart disease), Thalach (maximum heart rate achieved), ST depression (finding on an electrocardiogram, trace in 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 employing a variety of automated classification techniques, such as naive Bay Probabilistic Forest, and SVM, in identifying cardiovascular conditions 43% of all fatalities, according to the Global Burden of Ailment Schoolwork 2017. Unhealthy diet, smoke, excessive sugar, and being overweight or obese are all mutual risk features for heart disease 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.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 7, July 2023

DOI: 10.17148/IARJSET.2023.107105

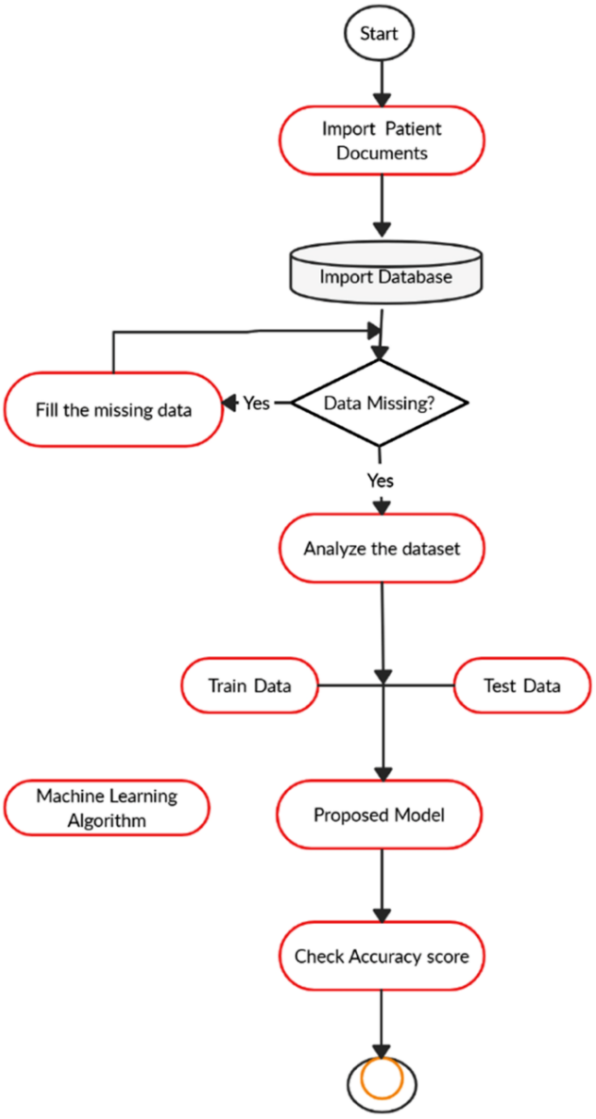


Fig. 1. Proposed Architecture

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 sentiment disease prediction system. Traditional approaches for envisaging sentiment 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 glowing as up-to-the-minute data pulling out procedures are 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 construction of fresh solutions.We use Reproduction Neural System (ANN) to envisage sentiment disease in this suggested method. The project "Heart Disease Prediction Using Artificial Neural Network (ANN)" intends to use ANNs to create a prediction model for the early judgment of heart disease.ANNs are cultured machine scholarship systems that 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 Cleveland Sentiment Ailment dataset, which is available on the UCI appliance scholarship storehouse / 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 envisage the prospect of sentiment ailment in patients with great accuracy.The trained model is then evaluated on a separate dataset for performance and accuracy. Finally, the project intends to create a user-

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 7, July 2023

DOI: 10.17148/IARJSET.2023.107105

friendly interface that will allow doctors and healthcare workers to enter patient data and obtain expected results.The suggested methodology has the likely to increase the accuracy and speed with which cardiac disease is diagnosed, allowing for earlier intervention and improved patient outcomes. This initiative has the potential to significantly affect public health by contributing to the expansion of AI-powered healthcare solutions.

II. LITERATURE REVIEW

Numerous attempts have been made in health care organizations to develop sickness prediction systems that make use of various data mining techniques and machine learning algorithms. Forecasters of Cardiac Disease Using a Complex Regression Model, which indicates that a mix of linear regression models is useful to determine heart disease risk, was proposed by K. Polaraju et al. [7]. A training data collection of 3000 instances with each of the aforementioned 13 different attributes is used for the job. The data set is divided into two halves, with 30% of the data used for qualification and 70% of the data used for training. It is clear from the statistics that the The descent method outperforms the other algorithms in relations of classification accuracy.

Marjia et al.[8] created A layered perceptual predictions model for heart illness utilizing WEKA software, thereby KStar, j48, SMO, and Bayes Net. SMO and Bayes Net surpass KStar, Multilayer Perception, and J48 techniques making use of cross-validation at k-folds when it comes to of accuracy from several criteria. The technologies' accuracy performance are still not sufficient. Consequently, the precision of the performance is increased further to provide better decisions for complaint judgment.

S. Seema et al. [9] focus on approaches for The algorithms known as Nave Bayes a decision-tree, Support Vector Machine (SVM), and Artificial Neural Network (ANN) are used to mine past health records for data and forecast chronic illness. In order to discover which predictive algorithms work better in terms of the precision rate, an assessment is done on them. SVM had the highest accuracy percentage of this trial, however Nave Baye is more accurate for diabetes has the highest accuracy.

Numerous techniques, including Naive Bayes, This step Trees, KNN, statistical regression, SVM, and ANN, were recommended by Ashok Kumar Dwivedi et al. [10]. Comparing graphical regression to other algorithms, it offers higher accuracy.

MeghaShahi et al. suggested the Heart Disease Identification A framework Using Data Mining Techniques [11]. Healthcare facilities employ WEKA instruments to perform sickness notification and to deliver high-quality treatment. Some of the algorithms used in the study included SVM, Nave Bayes, The connection rule, KNN, ANN, and Tree of Decisions work. The report suggested that SVM is more effective and accurate than other data pulling out techniques.

III. PREVIOUS SYSTEM

The current approach predicts cardiac utilizing several algorithms for methods to classify for illness, including the naive Bayes method, the the random forest, and SVM, and others, and compares their accuracy ratings.

The current heart ailment calculation arrangement employs Naive Bayes, Random Forest, and Funding Vector Machines (SVM). In machine learning, these systems are 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. While these set of rules have had considerable success, they do have significant limits that can distress their accurateness 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 with the use of the training set to associate their routine using their accuracy scores. Stacking Ensemble Learning will be employed later to progress the presentation of the underlying classification models.

Overfitting: Sometimes ensemble learning approaches overfit the training data, resulting in excellent accuracy on the working out data but poor routine on 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 sentiment ailment calculation can be difficult. This may impede physicians' and healthcare professionals' capacity to apply the model in decision-making procedures.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 7, July 2023

DOI: 10.17148/IARJSET.2023.107105

Data Bias: Ensemble learning methods are susceptible to data bias, which occurs when the model is accomplished on unbalanced datasets, resulting in biassed predictions. This can have an impact on the accuracy and dependability of the forecasts.

Ensemble learning approaches may be computationally costly, requiring large processing resources to Eurostar and optimise. This may check their efficacy 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.

Related Work:

In the suggested method, we use Machine Scholarship, specifically Artificial Neural Networks (ANN), to predict heart disease. The suggested technique entails creating an ANN prediction prototypical that can reliably identify people at peril of temperament virus. The method use ANNs to expect the fortuitous of heart illness using the prominent dataset Cleveland Heart illness, which is accessible on the UCI appliance scholarship repository / 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 can learn from preprocessed data and reliably forecast the likelihood of heart ailment in patients in the proposed system. An artificial neural system (ANN), often known as a neural network, is a mathematical model based on biological brain networks. The reproduction neural system is based on the study of the human brain. The human brain is a complex network of neurons. Axons, dendrites, and synapses are components of neurons. The planned ANN is made up of three layers: an input layer, a hidden layer, and an output layer. The suggested system is built with the Flask web framework, and the trained ANN model is deployed as a web application that can be edited by any 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 in its premature points, consenting for 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.

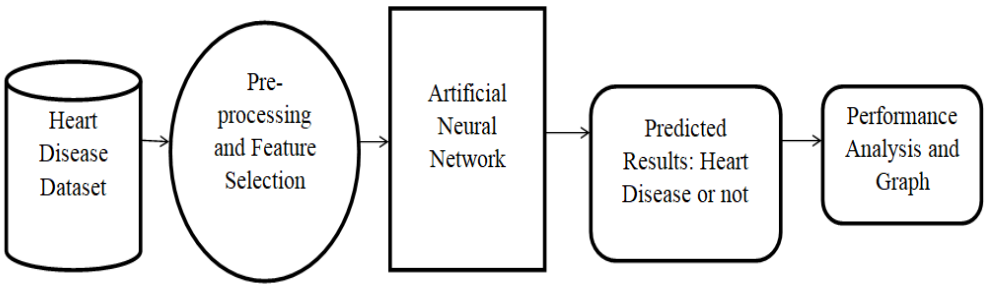
Speed: Because the ANN-based prediction model can process data fast, it can provide speedy findings that can be used for timely interventions and treatment.

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 more severe types of heart disease.

Improved Patient Outcomes: The ANN-based prediction model can progress easy-going conclusions and lower death rates by identifying people at risk of sentiment sickness and delivering early therapies.



IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 7, July 2023

DOI: 10.17148/IARJSET.2023.107105

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

IV. METHODOLOGY

Machine Learning:

Machine learning is a developing technique that allows computers to learn autonomously from historical data. Machine learning employs a variety of algorithms to construct mathematical models and make predictions based on past data or information. It is being utilised for a variety 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 as well as quite a few machine learning tactics such as supervised, unsupervised, and reinforcement learning. Regression and classification models, clustering approaches, hidden Markov models, and other sequential models will be covered.

In the actual world, we are enveloped by individuals who have the potential to acquire from their own breathes, as well as 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 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.

V. CONCLUSIONS

Major influencing elements for determining heart disease and numerous research efforts in predicting heart disease have been found and described in this study. It has been initiate that not all qualities are considered by every researcher. Few characteristics are deleted by a few researchers in direction to deliver more accuracy. We had a thorough argument on the major issues of various research studies for heart disease calculation that have yet to be solved. In the future, researchers should use an effective algorithm to incorporate all of the elements for identifying heart disease. As a result of the literature assessment, there is a need for combinational and more complicated models to progress the accurateness of forecasting the early beginning of cardiovascular illnesses.

The study presents a method for accurately predicting cardiac disease by combining support vector machines and artificial neural networks. Using the Cleveland Heart Disease database, we want to give guidance for training and testing 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 may be identified.

REFERENCES

1) Sharmila S et al., “Analysis of Heart Disease Prediction using Data Mining Techniques”, International Journal of Advanced Networking & Applications (IJANA), Volume: 08, Issue: 05, Pages: 93-95 (2017).
2) Chaitralii S. Dangaree et al., “Improved Study of heart disease Calculation system using Data Taking out rrangement Techniques”, International Journal of Computer Applications, Volume: 47, Pages: 44-48 (2012).
3)Animesh Hazra, 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 and Technology, ISSN 0973- 6107 Volume 10, Number 7 (2017) pp.2137-2159.
4) What is Predictive Data Mining?, 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)B.Venkatalakshmi, M.V Shivsankar, “Heart Disease Diagnosis Using Predictive Data mining”, International Journal of Innovative Research in Science,Engineering and Technology, Volume3, Special Issue 3, March 2014.

DOI: 10.17148/IARJSET.2023.107105

9)Sunil Ray, "Learn Naïve Bayes algorithm" and "Decision tree- Simplified", URL: www.analyticsvidhya.com, Retrieved on- 10.01.2018.

10)Kalaiselvi C, "Diagnosing of heart diseases using Average K- Nearest Neighbor Algorithm of Data Mining", 2016 InternationalConference on Computing for Sustainable Global Development (INDIACom), IEEE, Pages: 3099-3103 (2016).

11) 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).

12) Theresa Princy R, "Human Heart Disease Prediction System using Data Mining Techniques", International Conference on Circuit, Power and Computing Technologies [ICCPCT], IEEE (2016).

13)Vivekanandan T et al., "Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease", www.elsevier.com/locate/compbiomed, https://doi.org/10.1016/j.compbiomed, Pages: 125-136 (2017).

14)Veterinary Science (IOSR- JAVS), e- ISSN: 2337-2380, p-ISSN: 2337-2372, Volume: 4, Issue: 2, Pages: 61-64 (2013).

15) John Peter Tet al., "An Empirical study on Prediction of Heart Disease using Classification Data Mining Techniques", International Conference On Advances in Engineering, Science and Management (ICAESM -2012), IEEE, ISBN: 978-81- 909042-2-3,

