International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 733-737 www.ijaem.net ISSN: 2395-5252

A survey on heart disease prediction and detection Using ML Approach

[1]Asst.Prof, Sadhana B, [2]Student, Ankith, [3]Student, Jeevan, [4]Student, Pradeep, [5]Student, Sohan

Canara Engineering College Mangalore, Karnataka
Canara Engineering College Mangalore, Karnataka
Canara Engineering College Mangalore, Karnataka
Canara Engineering College Mangalore, Karnataka
Canara Engineering College Mangalore, Karnataka

Submitted: 15-12-2022 Accepted: 25-12-2022

ABSTRACT- ML is a useful tool in the healthcare industry since it can determine whether locomotor disorders, heart ailments, and other conditions will exist or not. This project uses ML methods including logistic regression, backward elimination, and REFCV to estimate a person's risk of acquiring heart disease using a readily available Kaggle dataset. After that, a confusion matrix and cross validation are used to assess the outcomes. Early diagnosis of cardiovascular conditions can assist patients in altering their lifestyles to reduce the risk of complications, which may represent a significant medical advance

KEYWORDS: Machine Learning, Convolutional Neural Network (CNN), Deep Learning, Support Vector Machine (SVM), Random Forest Algorithm (RFA).

I. INTRODUCTION

The research that is recommended in this paper largely focuses on various data mining methods used to forecast cardiac disease. The heart is the primary muscle in the human body. In essence, it regulates how blood moves through our body. Any cardiac condition can make other body parts more painful. Any illnesses that make it difficult for the heart to beat normally are categorized as heart diseases. Heart disease is one of the leading causes of death in the modern world. Heart disease can be brought on by living a sedentary lifestyle, smoking, drinking alcohol, and eating a lot of fat, which can raise blood pressure. According to the World Health Organization, heart disease claims the lives of more than 1 million individuals globally each year. The only means of stopping heart-related diseases are a healthy lifestyle and early detection The primary difficulty in modern healthcare is the provision of excellent services and quick, accurate diagnoses. Despite the fact that heart disorders are now the biggest cause of death worldwide, they are also the ones that are most easily managed and controlled.

The proper time of disease detection determines how accurately a disease will be managed overall. The proposed work aims to identify these heart conditions early in order to prevent negative outcomes. For analysis and knowledge extraction, records of a substantial collection of medical data gathered by medical practitioners are available. By using data mining techniques, it is possible to extract crucial and hidden information from the large amount of data that is now available. Discretised data make up the majority of the medical database. Using discrete data to make decisions is a difficult and tough undertaking.

ML, a branch of data mining, efficiently handles enormous, meticulously ordered datasets.

A wide range of disorders can be identified, detected, and predicted using machine learning in the medical sector.

II.LITERATURE SURVEY

The biggest cause of death worldwide, cardiovascular illness, is discussed in this study as a potential target for machine learning techniques.Cross-validation and backward elimination are two algorithms that are used to increase the accuracy of the prediction. However, the accuracy of these algorithms is still not satisfactory, so the paper suggests improving their performance in order to provide better decision-making in diagnosing diseases. The Recursive Feature Elimination (RFE) method is proposed, which involves fitting a model repeatedly and removing the weakest feature until the desired number of features is reached.

To effectively predict the risk levels of heart illnesses, computer-based system diagnostic tools are built on the boosting Support Vector Machine (SVM) algorithm. This study employed the Cleveland Clinic Foundation dataset, which contains 14 variables relating to patients' vital signs. Of these fourteen characteristics, thirteen are employed as predictor variables, and the final characteristic is used as the target or anticipated class. Boosting is an ensemble strategy that transforms weak learners into strong learners by effectively eliminating dataset biases for machine learning algorithms. The boosting strategy's goal is to improve prediction precision.

In order to increase prediction accuracy, this research suggests a method for the diagnosis of heart disease utilising ensemble classification algorithms. Individual classifiers were trained on separate sets of data, a training set and a test set and then their effectiveness was evaluated using the test set. A significant boost in prediction accuracy followed the implementation of feature selection to further improve the process's performance. Feature selection was then implemented to further enhance the performance of the process, resulting in a significant improvement in prediction accuracy. For poor classifiers, the highest accuracy gains of 7% was attained.

"Heart Disease Prediction Using Classification with Different Decision Tree Techniques" proposes a system that uses data mining techniques and decision tree techniques to predict heart disease. ID3 and C4.5 are two decision tree models used in this system. ID3 stands for Iterative Dichotomiser 3, and is an extension of C4.5. Both approaches can generate classifiers that are described as either rulesets or decision trees. The quantity of information gain is measured by entropy, and the method utilised in this system chooses the splitting attribute that minimises entropy while maximising information gain.

This study suggests using a trained Recurrent Fuzzy Neural Network (RFNN) built using a Genetic Algorithm (GA) to identify cardiac problems. The network tested with 97.78% accuracy, according to the results. Root Mean Square Error (RMSE), sensitivity (recall), specificity, accuracy, F-score, and chance of misclassification error were used to further evaluate the network (PME). The weights and biases were encoded as 64-bit long genes, and the RFNN had 13 inputs, 7 shrouded neurons, and 1 output neuron. GA was employed with a population size, a multipoint crossover chance of 0.25, and a mutation probability of 0.05. The GA-based trained RFNN technique was found to have satisfactory performance.

In order to identify the best network design for heart disease prediction, this research suggests a neuro-genetic method.

The combination of the algorithm and classifier achieves a 96.25% accuracy rate. Metrics are proposed to provide insight into reliable factors. The technique determines if a person has cardiac disease based on elements in their medical record using a CNN-based uni-modal disease risk detection algorithm.

The neural network algorithm is used to identify risk factors related to heart disease, by using both structured and unstructured textual data. The Neuro-Genetic approach uses a multi-objective function for the selection process of the feature subset through an efficient GA.

This paper proposes using logistic regression and naïve baiyes algorithms to diagnose chronic ischemic heart diseases. This study aims to properly estimate a person's risk of acquiring heart disease. Data for the study was gathered via an open dataset platform (Kaggle), and any missing data was addressed by analysis utilising traditional statistical measures (Mean, Median, and Mode). Following data cleaning, analysis of data properties and their intended use was carried out. There were a total of 165 instances of the gender attribute with a danger rate of 1 and 207 instances of the gender attribute with a risk rate of 0.

This paper proposes a system for heart disease prediction that utilizes a combined approach of two Multilayer Perceptron Neural Networks (MLP). The system's performance is assessed using the 303 instances and 76 attributes of the Cleveland Hospital Heart Disease Dataset from the UCI Machine Repository, with the majority of the research concentrating on 14 of those attributes. To create a reliable cardiac disease prediction model, the Artificial Neural Network (ANN) method is frequently utilised, and MATLAB is the tool of choice.

In addition, the ECG data from the PhysioNet ECG Database of diseased patients is used in conjunction with the clinical features.

In this study, a hybrid classification system for the diagnosis of heart illness is proposed. It consists of an ensemble classifier and a feature selection subsystem called ReliefF. According to a jackknife cross-validation scheme, this system outperformed previously published classification algorithms with a maximum classification accuracy of 92.59%. While Rough Set (RS) theory is a mathematical strategy that resolves ambiguity and uncertainty, ReliefF is a widely applied and successful feature selection algorithm.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 733-737 www.ijaem.net ISSN: 2395-5252

In order to automatically identify cardiac illness, this work describes a system that combines an Artificial Immune Recognition System (AIRS) with a fuzzy resource allocation mechanism and k-nn. The UCI ML Database provided the dataset for the study. The system's accuracy of 87% is promising and greater than the accuracy of other applications for this topic in the literature. The high-affinity ARBs can function more efficiently in AIRS due to the competition for resources. With this resource distribution approach, the antigen class ARBs receive half of the resources and the other classes receive the other half. By multiplying the stimulation rate by the clonal rate, the distribution of resources is calculated.

A system that compares the accuracy of various data mining classification schemes and their combinations through Ensemble ML Techniques for predicting heart disease is proposed in the paper "A Comprehensive Investigation and Comparison of ML Techniques in the Domain of Heart Disease." This study makes use of the 303-instance Cleveland dataset for cardiac disorders. The data is split into training and testing datasets using 10-Fold Cross-Validation due to the small amount of samples. These include Decision Tree (DT), Naive Bayes (NB), Multilayer Perceptron (MLP), K-Nearest Neighbor (K-NN), Single Conjunctive Rule Learner (SCRL), Radial Basis Function (RBF), and Support Vector Machine (SVM). When working with small data sets for training and testing sets, K-Fold Cross Validation is the recommended data portioning strategy. According to the number of folds, the dataset is split into two portions in this method for training and testing in each iteration.

The use of Moth-Flame Optimization (MFO) for the early diagnosis of heart problems is examined in this research. To generate a single, best-fitting predictive model, the proposed method merges four categorization models utilising ensemble stacking and logistic regression. The classification system's accuracy is increased while its processing time is decreased through the use of feature selection techniques. MFO is an optimization method based on moths' propensity to face the Moon when flying far distances.

Support Vector Machines (SVMs) are also utilized to tackle classification and regression issues.

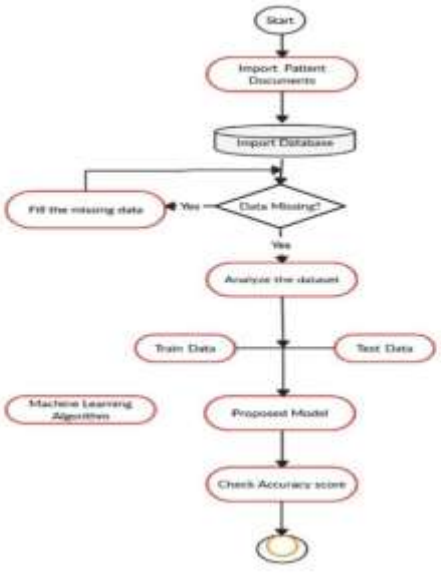
SL.NO	Title of the paper	Existing System	Methodology/ Algorithm	Drawback
1	Heart disease prediction using machine learning algorithms	Determining the significant risk factors based on medical dataset which may lead to heart disease	Recursive feature Elimination using Cross-validation and Backward Elimination	Here if once a variable is eliminated from a model it is not re-entered again
2	Computer-Aided Diagnostics of Heart Disease Risk Prediction Using Boosting Support Vector Machine	The Long-term Recurrent Convolution Network (LRCN) was used for temporal analysis of the cropped frames of eye blinking.	Support Vector Machine	It does not execute very well when the data set has more sound i.e. target classes are overlapping.
3	Improve the accuracy of prediction of heart disease risk based on ensemble classification techniques	The machine learns patterns from the existing dataset and then applies to an unknown dataset inorder to predict the outcome	Ensemble algorithms	This method is less interpretable and more time consuming
4	Heart disease prediction using classification with different decision tree techniques	Decision trees for identifying various ways of splitting datasets into segments	Iteartive Dichomotizer(ID3), C4.5 and C5.0 classifier	Overfitting or high varience which means that it learns the dataset so well that it fails to generalize on new data

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 733-737 www.ijaem.net ISSN: 2395-5252

6	Heart disease prediction using Neuro-Genetic Algorithm and CNN-MDRP classifier	Identifies whether an individual undergoes critical symptoms of heart disease based on the aspects in medical record	Neuro-Genetic Algorithm , and a Convolutional Neural Network (CNN).	Formulating a differentiable loss function, that follows the proposed signal processing step is not easy
7	Diagnosis of Chronic Ischemic Heart Diseases using Machine Learning Algorithm	Predicts precise diagnosis of chronic dataset where the system is trained using a training dataset and then tested using test dataset	Logistic regression and Naïve Baiyes algorithm	Fails to predict a continuous outcome and may not be accurate if the sample spaces is too small
8	Analysis of heart disease prediction using Artificial Neural Network	Existing system is mainly focused on clinical attributes only	Artificial Neural Network (ANN) which is a learning algorithm	Neural network training requires a lot of data
9	A hybrid Classification System for Heart Disease Diagnosis based on the RFRS method	To aid the diagnosis of heart disease using a hybrid classification system based on the RelieFF and Rough Set (RFRS) method	ReliefF algorithm and C4.5 classifier	Don't handle data that belong to quantitative
10	Automatic detection of heart disease using an artificial immune recognition system with fuzzy resource allocation mechanism and k-nn	Focuses on executing a supervised learning algorithm (AIRS) for classification	Fuzzy resource allocation and K-nn algorithm	With large data prediction rate may be slow
11	A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease	To investigate and compare the accuracy of different data mining techniques	MultiLayer Perceptron	The output values of a perceptron can take on only one of 2 values due to the hard-limit transfer function
12	Moth-Flame Optimization for Early Prediction of Heart Diseases	Provide a unique method for finding essential traits employing machine learning approaches in this paper, which enhances the effectiveness of identifying heart diseases	Moth Flame Optimization(MFO) algorithm	Algorithm is highly sensitive to parameter and has slow convergence

III.PROPOSED SYSTEM

Here is a summary of the a whole deep learning research methods, including how the neural network models is filled into the initiative and how it would be processed. This data is communicated in both text and schematic diagram format. Abbreviations and Acronyms



Schematic View for Prototype Recommendation

The Cleveland heart dataset, which has been acquired from UCI's data source, is used to evaluate the study's validity. Each documentation in this dataset has thirteen clinical characteristics, including age, physical intimacy, the form of chest trauma caused, resting bp, lipids, breakingthe fast glucose levels, laying Electrocardiogram(ECG), highest pulse rate, workout heart problems, old high point, gradient. Actual data contains vast quantities of incorrect and incomplete information. When gathering data, loudness and incompleteness are usually present. Significance of this research from one type to the other in order to improve it's own legibility is known as transition. In this plan, the file is divided randomly into two frames: preparation and study. A dataset that has been divided into a training dataset and a test dataset uses the significant data for training with testing. Collection And analysis chooses the information at random to make sure that train and test sets are equivalent. Visitors can diminish the impact of inconsistent data and give a deeper understanding of the model's character traits using comparable data to train and study.

IV.CONCLUSION

It is important to create a structure which can anticipate heart conditions clearly and efficiently which can lead to rise in deaths due to heart illnesses. Trying to find a most efficient ML algorithm for cardiac disease detection was the original study main driver. Utilizing data from the Uci machine learning, this report examines the accuracy rate of the Decision Tree, Logistic Regression, Random Forest, and Naïve Bayesian methodologies for predicting the presence of heart disease. Upcoming projects may be deemed to be managed in cooperation with clinical personnel to include specific elements that may impact the product's judgement capabilities. Trying to test the performance of the proposed framework using various data sources from other source materials may also be beneficial.

REFERENCES

[1] T. Nagamani, S. Logeswari, Gomathy," Heart Disease Prediction using Data Mining with Mapreduce Algorithm", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-3, January 2019.

[2] Fahd Saleh Alotaibi," Implementation of Machine Learning Model to Predict Heart Failure Disease", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 6, 2019.

[3] Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin,"Design and Implementation Heart Disease Prediction Using Naives Bayesian",

[4] Theresa Princy R, J. Thomas,'Human heart Disease Prediction System using Data Mining Techniques', International Conference on Circuit Power and Computing Technologies, Bangalore,2016.

[5] Nagaraj M Lutimath, Chethan Basavaraj S Pol.,'Prediction of Heart Disease using Machine Learning', International journal Of Recent Technology and Engineering,8, (2S10), pp 474-477, 2019

[6] Jafar Alzubi, Anand Nayyar, Akshi Kumar. "Machine Learning from Theory to Algorithms: An Overview", Journal of Physics: Conference Series, 2019.

[7] Fajr Ibrahem Alarsan., and Mamoon Younes 'Analysis and classification of heart diseases using heartbeat features and machine learning algorithms',Journal Of Big Data,2019;6:81.

[8] Himanshu Sharma and M A Rizvi, "Prediction of Heart Disease using Machine Learning Algorithms: A Survey" International Journal on Recent and Innovation Trends in Computing and Communication Volume: 5 Issue: 8 , IJRITCC August 2019.

[9] M. Nikhil Kumar, K. V. S. Koushik, K. Deepak, "Prediction of Heart Diseases Using Data Mining and Machine Learning Algorithms and Tools" International Journal of Scientific Research in Computer Science, Engineering and Information Technology ,IJSRCSEIT 2021

[10] Amandeep Kaur and Jyoti Arora,"Heart Diseases Prediction using Data Machine Techniques: A survey" International Journal of Advanced Research in Computer Science , IJARCS 2017-2021.

[11] Pahulpreet Singh Kohli and Shriya Arora, "Application of Machine Learning in Diseases Prediction", 4th International Conference on Computing Communication And Automation(ICCCA), 2020.