ORIGINAL RESEARCH ARTICLE

The Effectiveness of Machine Learning Systems' Accuracy in Predicting Heart Stroke Using Socio-Demographic and Risk Factors - A Comparative Analysis of Various Models

Nihar Ranjan Panda[1], Kamal Lochan Mahanta[2], Jitendra Kumar Pati[3], Ruchi Bhuyan[4*], Soumya Subhashree Satapathy[5]

[1]CV Raman Global University, Bhubaneswar, Odisha, India
[2]CV Raman Global University, Bhubaneswar, Odisha, India
[3]Kiit International School, KIIT University, Bhubaneswar, Odisha, India
[4]IMS and SUM Hospital, SOA deemed to be University, Bhubaneswar, Odisha, India
[5]Center for Biotechnology, School of Pharmaceutical Science, SOA deemed to be University, Bhubaneswar, Odisha, India

DOI: 10.55489/njcm.140620233026

ABSTRACT

Background: Cardiologists can more appropriately classify patients' cardiovascular diseases by executing accurate diagnoses and prognoses, enabling them to administer the most appropriate care. Due to machine learning's ability to identify patterns in data, its applications in the medical sector have grown. Diagnosticians can avoid making mistakes by classifying the incidence of cardiovascular illness using machine learning. To lower the fatality rate brought on by cardiovascular disorders, our research developed a model that can correctly forecast these conditions.

Methods: This study emphasized a model that can correctly forecast cardiovascular illnesses to lower the death rate brought on by these conditions. We deployed four well-known classification machine learning algorithms like K nearest Neighbour, Logistic Regression, Artificial Neural network, and Decision tree.

Results: The proposed models were evaluated by their performance matrices. However logistic regression performed high accuracy concerning AUC (0.955) 95% CI (0.872-0.965) followed by the artificial neural network. AUC (0.864) 95% CI (0.826-0.912).

Conclusion: Individuals' risk of having a cardiac event may be predicted using machine learning, and those who are most at risk can be identified. Predictive models may be developed via machine learning to pinpoint those who have a high chance of suffering a heart attack

Keywords: Machine learning, cardiovascular disease, neural network, Prediction

ARTICLE INFO

Financial Support: None declared
Conflict of Interest: None declared
Received: 28-04-2023, **Accepted**: 19-05-2023, **Published**: 01-06-2023
***Correspondence:** Ruchi Bhuyan (Email: ruchibhuyan@soa.ac.in)

How to cite this article:
Panda NR, Mahanta KL, Pati JK, Bhuyan R, Satapathy SS. The Effectiveness of Machine Learning Systems' Accuracy in Predicting Heart Stroke Using Socio-Demographic and Risk Factors - A Comparative Analysis of Various Models. Natl J Community Med 2023;14(6):371-378. DOI: 10.55489/njcm.140620233026

www.njcmindia.com | pISSN09763325 | eISSN22296816 | Published by Medsci Publications

INTRODUCTION

Cardiovascular disorders affecting the heart and blood vessels are becoming a global burden, accounting for more than 70% of mortality.[1] According to the WHO, CVD claims up to 17.9 million deaths per year. CVD constitutes an array of disorders, including angina, stroke, heart failure, carditis, heart attack, rheumatic heart diseases, venous thrombosis, peripheral artery disease, and numerous other conditions.[2] CVD-associated risk factors include hypertension, smoking, hyperlipidemia, obesity, stress, a poor diet, and a family history of CVD.[3] Electrocardiograms, echocardiograms, magnetocardiography, and magnetic resonance imaging are all used to diagnose CVD. There are, without a doubt, various diagnostic options; nonetheless, their limitations are a nuisance. The ECG cannot offer a conclusive diagnosis of congestive heart failure. One of the most significant drawbacks of cardiac echocardiography is that it does not reveal any blockages or coronary arteries. Although magnetocardiography produces high-quality signals, it is a time-consuming procedure.[4] Furthermore, these diagnostic options are sometimes excessively expensive and impracticable for patients in middle and low-income countries.[5]

A stroke is a medical crisis instigated by an intrusion in blood flow to the brain that ends up in cell death and loss of brain function. Strokes have been classified as either ischemic or hemorrhagic. Both forms of stroke can cause considerable brain damage and result in a variety of physical and cognitive symptoms.[6] Stroke symptoms often appear suddenly, within seconds to minutes, and do not proceed further in the majority of instances. Dysarthria, aphasia, ptosis, and altered taste and smell are some of the indications. The risk factors of heart stroke embrace heart disease, high RBC count, high level of cholesterol, high blood pressure, diabetes, unhealthy diet, and second-hand smoking. Delayed medical presentation and inability to comply with medications are some of the primary issues that must be resolved.[7]

Over time, an assortment of clinical procedures has been designed to assist in determining the existence of stroke. Whilst these procedures can help with the first triage of acute neurological patients, they are unable to match both the specificity and sensitivity of an imaging evaluation nor is there a clinical test that can extricate among ischemic and haemorrhagic stroke.[8] Heart stroke can be diagnosed using a CT scan, MRI, and electrocardiography. During a CT scan, a patient is exposed to ionizing radiation, which can cause long-term harm and potentially raise the possibility of cancer. According to the WHO, the worldwide incidence of CVD deaths is expected to upsurge to 23.6 million by 2030, with cardiovascular disease and stroke being the main culprits.[9]

Machine learning algorithms can detect early warning signals of heart disease and stroke, allowing for earlier prevention and therapy.[10] This can scan massive amounts of data and uncover patterns that specialists may overlook, potentially extending lives while improving outcomes. This may give rise to further precise diagnoses and treatment recommendations. Moreover, ML algorithms can assess patient data and recommend individualized treatment strategies based on criteria such as age, gender, medical history, and lifestyle.[11]

METHODOLOGY

The data set used for this research is obtained from the Kaggle website i.e., https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset, which is openly available 5110 observations with 11 characteristics make up the data.[12] Kaggle which is a data-sharing website provides authentic and reliable secondary data for data scientists and researchers for research purposes. The particular data consists of 10 features which are described in Table 1. The outcome of the research is heating stroke which is a binary classification, yes (1), No (0). A k-fold cross-validation technique was used in machine learning techniques.

Proposed models

When speaking of a supervised learning issue in the setting of machine learning, a classification problem is one where the objective is to predict a categorical label or class variable for a given collection of features or input variables.

Table 1: Description of features and response levels

Feature	Description	Levels
Gender	Gender of the patient	Male (0), Female (1)
Work type	Description of work type	Children (0), Self-employed (1), Private (2), Never worked (3), Govt job (4)
Heart Disease	Whether associated with any heart disease	Yes (1), No (0)
Ever Married	The person is married or not	Yes (1), No (0)
Residence	Residence type	Rural (0), Urban (1)
Hypertension	Whether associated with hypertension	Yes (1), No (0)
Smoking status	Smoking status of the individuals	Formerly smoke (0,) Never smoke (1), Smokes (2), Unknown (3)
BMI	Body mass index of the individual	Mean (28.9)
age	Age of individual	Mean (43.2)
Avg glucose level	An average glucose level of the individuals.	Mean (106)

Table 2: Socio-demographic and risk factors associated with heart stroke

Variables	Stroke (%)	No stroke (%)	p-value
Total	249(4.9)	4861(95.1)	
Gender			
Male	108(43.4)	2007(41.3)	0.79
Female	141(56.6)	2853(58.7)	
work type			
Children	2(0.8)	685(14.1)	<0.001
Self-employed	65(26.1)	754(15.5)	
Private	149(59.8)	2776(57.1)	
Never worked	0(0)	22(0.5)	
Heart disease			
Yes	47(18.9)	229(4.7)	<0.001
No	202(81.1)	4632(95.3)	
Ever married			
Yes	220(88.4)	3133(64.5)	<0.001
No	29(11.6)	1728(35.5)	
Residence			
Rural	114(45.8)	2400(49.4)	0.269
Urban	135(54.2)	2461(50.6)	
Hypertension			
Yes	66(26.5)	432(8.9)	<0.001
No	183(73.5)	4429(91.1)	
Smoking Status			
Formerly smoke	70(28.1)	815(16.8)	<0.001
Never smoke	90(36.1)	1802(37.1)	
Smokes	42(16.9)	747(15.4)	
Unknown	47(18.9)	1497(30.8)	
BMI	30.4+5.8	28.9+7.7	0.002
Age	67.7+12.7	42+22.3	<0.001
Avg glucose level	133+61.9	105+43.8	<0.001

The class variable is referred to as the dependent variable or the target variable, whilst the input variables may be referred to as predictors or independent variables. The objective of a classification challenge is to discover a model or algorithm that can precisely predict the class variable for brand-new, undiscovered data points. Usually, a labelled dataset with known class labels for each data point is used to train the model. This labelled data is used by the model to discover patterns and connections between the predictors and the class variable. Decision trees, random forests, support vector machines (SVM), logistic regression, k-nearest neighbours (KNN), and neural networks are just a few of the techniques used in machine learning that may be employed for classification challenges. The particular issue and data, as well as the required level of accuracy, interpretability, and computing efficiency, all influence the algorithm's selection.

The dataset has been divided into a training set (80%) and a testing (20%). All computational and machine learning algorithms were employed in R version 4.3.0. The training dataset is employed to train a model, and the testing dataset is utilized to assess the model's performance. The effectiveness of multiple classifiers, K nearest Neighbour, Logistic Regression, Artificial Neural network, and Decision tree. has been assessed using the dataset. The efficacy of each classifier is then assessed using its ratings for recall, recall precision, accuracy, and F-measure.

K Nearest Neighbour Classification

A supervised machine learning approach for regression and classification analysis is the k closest neighbour (KNN). It is a non-parametric method; hence it makes no assumptions about the distribution of the data at its core. Instead, it memorizes the full training dataset and applies it to forecast data from fresh, unobserved bits of data. Identifying the K data points in the training set that are closest to the new data point and classifying the new point based on the majority class of those K neighbours is the fundamental concept underpinning KNN classification.[13,14] Although different distance metrics can be utilized, Euclidean distance is typically used to compute the distance between data points. Binary and multiclass classification issues can both be solved with KNN classification. Although it is a straightforward and understandable approach, as the amount of the dataset increases, it may become computationally expensive. KNN classification has the benefit of being a lazy learning algorithm, which eliminates the need for training time. Instead, predictions for fresh data points are made using the full training dataset. This implies, however, that prediction times can be lengthy, particularly for sizable datasets. A few KNN classification hyperparameters, including the value of K, the distance metric employed, and the procedure for identifying the majority class, can be tweaked to enhance performance. KNN classification is a flexible and reliable technique that may be applied to an extensive variety of classification issues.[15] However, to attain the greatest achievement, it is crucial to carefully choose the value of K and the distance metric.

Logistic Regression

Machine learning employs the statistical approach of logistic regression to solve categorizing issues.[16,17] When the target variable is categorical, which means it can only accept a finite number of values, this kind of supervised learning approach is employed. Discovering a relationship between the input features and the likelihood that the variable of interest will take a particular value is the aim of logistic regression. By estimating the parameters of a logistic function, which converts the input features into the likelihood of the target variable, this is accomplished. To minimize a cost function, such as the cross-entropy loss, which assesses the gap between the predicted probability and the actual target values, the logistic regression algorithm iteratively adjusts the parameter values. By computing the likelihood that the target variable will take a particular value based on the input features, the model may be used to make predictions on fresh data once the parameters have been evaluated.

Neural Network

Neural network algorithms are frequently employed for problem-solving in classification.[18,19] ANNs are made up of interconnected nodes (sometimes referred to as neurons) organized in layers and are de-

signed after the composition and operation of the human brain. An ANN's objective in a classification challenge is to figure out how to transform the input features into the output class labels.[20] Each data point in the labeled dataset used to train the network has a label for a particular target class. A loss function, which gauges the discrepancy between predicted and actual class labels, is minimized by the network during training by adjusting the weights and biases of the neurons in each layer. An input layer, one or more hidden layers, and an output layer are the common components of an ANN's architecture for categorization.[21,22] The input layer gets the features from the input layer, which is subsequently sent to the output layer via the hidden layers. Based on the input features, the output layer generates the expected class labels. Feedforward neural networks, convolution neural networks (CNNs), and recurrent neural networks (RNNs) are some of the ANN types that can be employed for classification. The most basic kind of ANN called a feed-forward neural network, is made up of many layers of interconnected neurons. CNNs use filters to extract features from the input images and are made to perform image recognition tasks. RNNs can account for the temporal dependencies between input features and are utilized for sequential data. In general, ANNs are an effective tool for classification problems because they can handle enormous quantities of data and learn complex non-linear correlations between the input variables and the output class labels

Decision tree

A machine learning approach known as a decision tree is employed for classification and regression issues.[23,24] It functions by creating a tree-like model that can be used for prediction by recursively splitting the data according to the values of the input characteristics.[25] Because they are simple to understand and can capture intricate non-linear correlations between the input data and the goal values, decision trees are widely used. They may be utilized for classification and regression issues and can handle both categorical and numerical input information. However, if the training data is not sufficiently regularized, decision trees may be sensitive to the selection of the splitting criterion and may overfit the data.[26] Decision trees can perform better and have less variation when using ensemble approaches like random forests and gradient boosting.

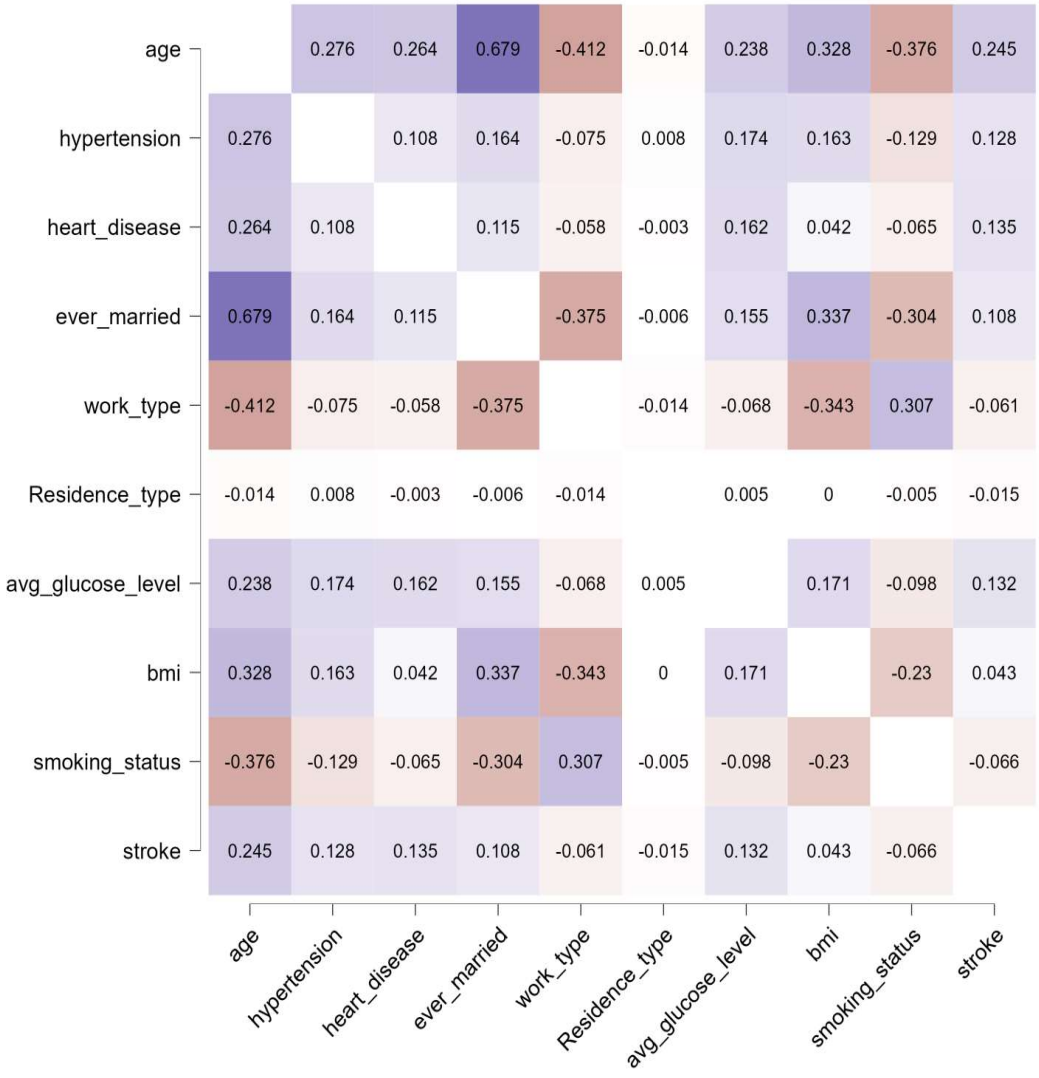

Figure 1: Correlation heatmap using all the features

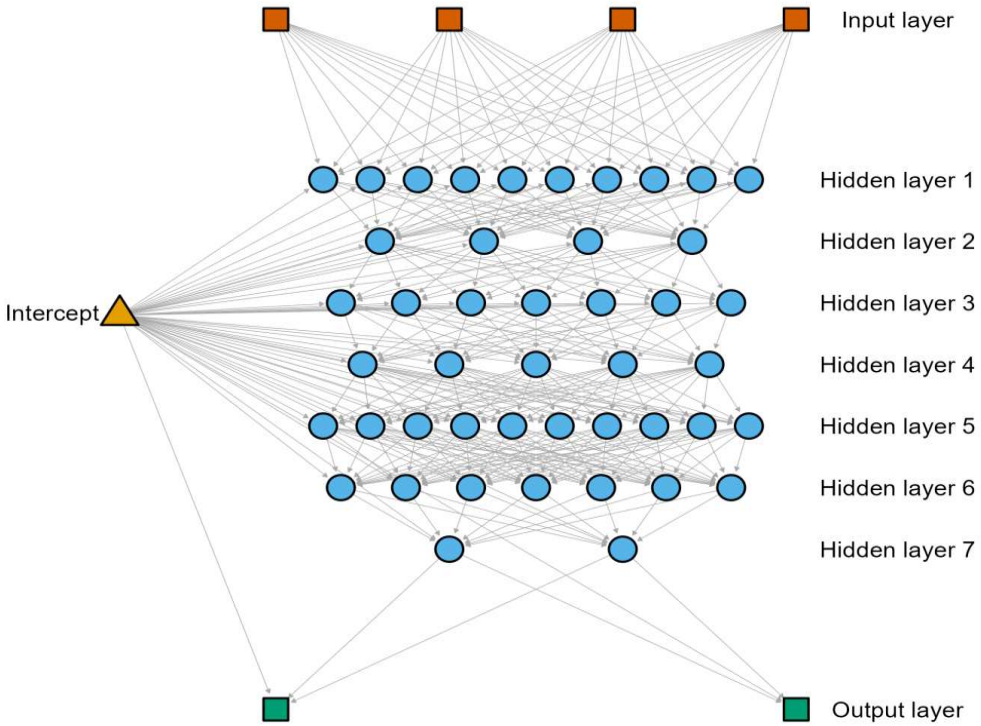

Figure 2: Schematic diagram of artificial neural network for predicting heart stroke

RESULTS

This study encompassed indicators of performance like as precision, recall, accuracy, F1 score, and area under the ROC curve. The dataset has been split into two halves, with 20% of the data used to test the model and 80% of the data used to train it. When assessing a classification model's efficacy, standard performance indicators include accuracy, precision, recall, F1 score, and AUC. The proportion of accurate predictions a model makes compared to all other forecasts is known as accuracy. It is calculated as (TP + TN) / (TP + TN + FP + FN), where TP is the total number of true positives, TN is the total number of true negatives, FP is the total number of false positives, and FN is the total number of false negatives, all of which were incorrectly predicted.

The fraction of accurate positive forecasts among all positive predictions is known as precision. To determine it, divide the total number of true positives by the total number of false positives, or TP / (TP + FP). The accuracy of the model's positive case identification is measured.

The recall is the proportion of correctly predicted positive outcomes among all instances of positive outcomes. TP / (TP + FN), where TP is the total number of true positives and FN is the total number of false negatives. Recall gauges a model's capacity to identify positive cases. The harmonic mean of recall and accuracy is the F1 score. As 2 * (precision * recall) / (precision + recall), it is computed. When the classes are unbalanced, the F1 score serves as a helpful statistic by balancing accuracy and recall. The statistic known as AUC (Area Under the ROC Curve) indicates the degree to which a model can distinguish between positive and negative situations. The true positive rate (recall) vs false positive rate (1 - specificity) for various categorization thresholds is plotted on the ROC curve. Higher numbers denote greater performance. The AUC stands for the area under the ROC curve and runs from 0 to 1. AUC is helpful when the cost of false positives and false negatives varies or when the classes are unbalanced.

The proposed ML Techniques deployed in this research produce the performance matrices. The accuracy, Precision, recall, F1 score, and AUC of the KNN model were 0.946, 0.895, 0.946.0.92, and 0.648 respectively. Similarly, logistic regression produces an accuracy of 0.955 with an AUC value of 0.864. The artificial neural network model performed an accuracy (0.955), Precision (0.903), recall (0.95), F1 score (0.926), and AUC (0.824) and the Decision tree produces an accuracy (0.956), Precision (0.931), recall (0.956), F1 score (0.937), and AUC (0.806). So, our research observed that Logistic regression performs better as compared to the other three ML techniques concerning AUC value.

Table 3: Performance matrices of all the proposed models

	Accuracy	Precision	Recall	F1 score	AUC
K-NN	0.946	0.895	0.946	0.92	0.648
LR	0.955	0.912	0.955	0.933	0.864
ANN	0.95	0.903	0.95	0.926	0.824
DT	0.956	0.931	0.956	0.937	0.806

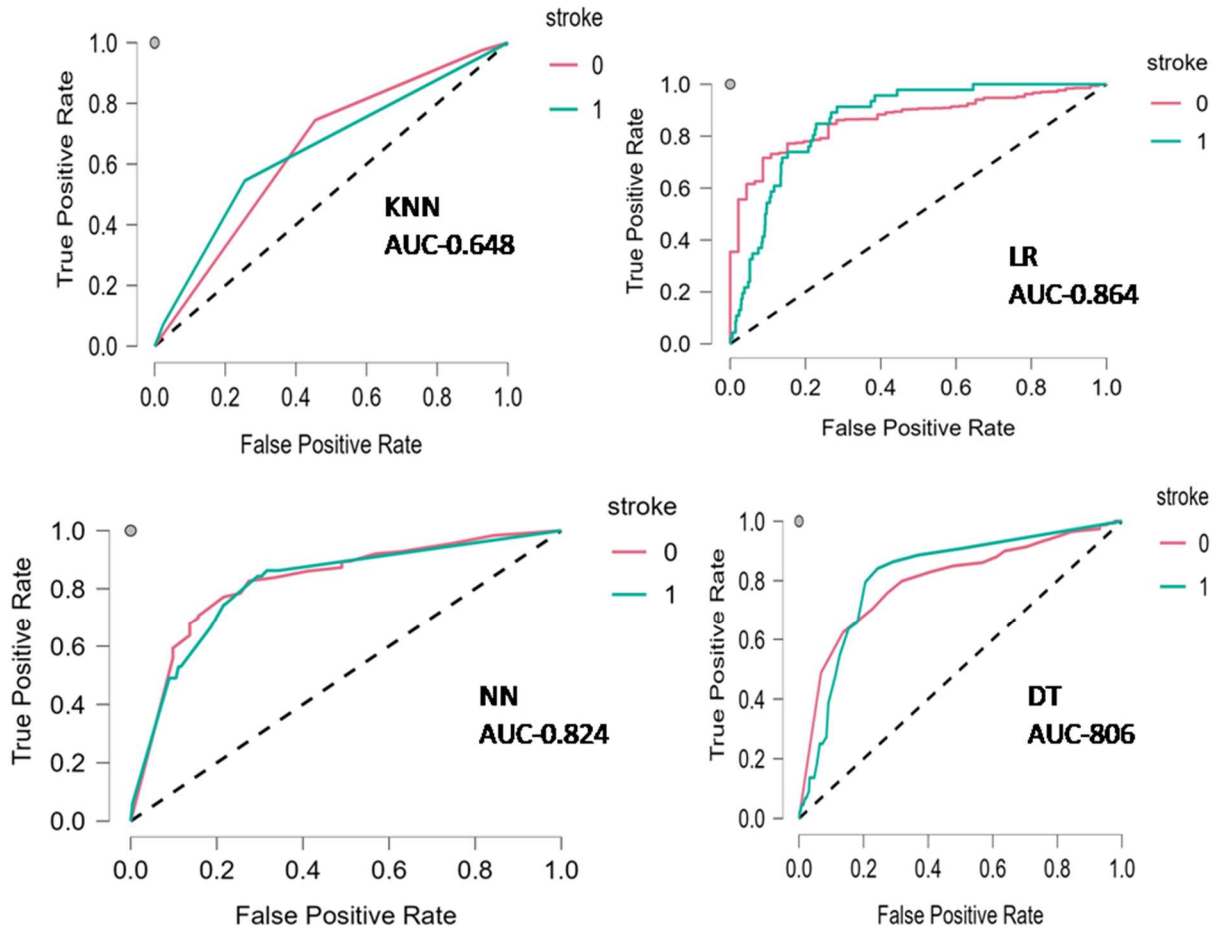

Figure 3: Area under the curve of all proposed models

Discussion

The use of ML techniques is showing efficacy in many different kinds of healthcare applications, most notably cardiovascular. Researchers have the chance to design and test new algorithms to identify risk factors and early indicators of heart ailments, which are still among the top causes of fatalities in developing countries. These approaches provide promising prospects for the early detection as well as mitigation of heart disease.[27] A survey of the literature reveals multiple researchers all through the globe use a wide array of ML approaches for predicting diseases.[28]

Drod et al. (2022) used ML approaches for determining substantial risk variables for CVD in individuals with metabolic-associated fatty liver diseases (MAFLD). They studied blood biochemistry and subclinical atherosclerosis data from 191 MAFLD patients and came up with a machine-learning model that could recognize those at elevated risks of CVD. The research investigation discovered that hypercholesterolemia, plaque scores, and diabetes duration were the most indispensable indicators of risk for CVD in MAFLD sufferers. With an AUC of 0.87, the ML method distinguished high- and low-risk patients well. The study reveals that utilizing basic patient criteria, ML methods can be useful in finding MAFLD individuals with extensive CVD.[29]

Rajathi and Radhamani (2016) used a combination of KNN and ACO algorithms to generate a model for predicting coronary heart disease (CHD). They used four ML algorithms to test the model's performance and attained an accuracy of 70.26%. The study reveals that combining KNN with ACO can be useful in predicting CHD. Further study, however, must be conducted to increase the model's precision.[30]

Shah et al. (2020) set out with the goal of developing a model for envisaging CVD using machine learning approaches. They used information from the Cleveland heart disease dataset, which had 303 occurrences besides 17 characteristics. To create the model, the scientists used supervised classification approaches such as random forests, naive Bayes, decision trees, and KNN. The KNN model has the utmost level of exactitude, at 90.8%, according to the data. The analysis emphasizes the probable application of machine learning in envisaging cardiovascular illness and accentuates the need of choosing proper models and methodologies to produce the best outcomes.[31]

Alotalibi's (2019) study sought to evaluate the use of ML procedures for forecasting heart failure disease. To construct prediction models, the researchers used a dataset from the Cleveland Clinic Foundation and utilized several ML methods for instance support

vector machine (SVM), logistic regression, naive Bayes, random forest, and decision trees. During the model-building procedure, a 10-fold cross-validation strategy was applied. The decision tree method has the best accuracy rate (93.19%), followed by the SVM algorithm (92.30%). The study stresses the decision tree algorithm as a feasible choice to consider in the subsequent study and underlines the impending ML approaches as a useful tool for forecasting cardiac maladies.[32]

Conclusion

Machine learning can forecast a person's likelihood of experiencing a cardiac episode and identify individuals who are most susceptible to it. To identify those who have a high risk of having a heart attack, predictive models can be created using machine learning. These models can evaluate several parameters, such as age, gender, family history, medical history, lifestyle, and other risk factors, to predict the chance of suffering a heart attack. Based on the results of the prediction model, patients may receive tailored recommendations for reducing their likelihood of having a cardiac event. A balanced diet, frequent exercise, regulating your body weight, and quitting smoking are just a few examples of lifestyle changes that may be advised. Using machine learning, it is possible to detect and keep track of people's health status and notify them if anything changes that would increase their risk of suffering a heart attack. For more advanced use of these methodologies, health data may be collected and analyzed by wearable technology, smartphone applications, and other technologies.

Abbreviation

KNN: k nearest neighbour
LR: Logistic regression
DT: Decision tree
ANN: Artificial neural network
CVD: Cardiovascular disease

References

1. Mukherjee D, Patil CG. Epidemiology and the global burden of stroke. World neurosurgery. 2011 Dec 1;76(6):S85-90.
2. Benziger CP, Roth GA, Moran AE. The global burden of disease study and the preventable burden of NCD. Global heart. 2016 Dec 1;11(4):393-7.
3. Tran DM, Lekhak N, Gutierrez K, Moonie S. Risk factors associated with cardiovascular disease among adult Nevadans. PloS one. 2021 Feb 17;16(2):e0247105.
4. Jahmunah V, Oh SL, Wei JK, Ciaccio EJ, Chua K, San TR, Acharya UR. Computer-aided diagnosis of congestive heart failure using ECG signals–a review. Physica Medica. 2019 Jun 1;62:95-104.
5. Bhatt CM, Patel P, Ghetia T, Mazzeo PL. Effective Heart Disease Prediction Using Machine Learning Techniques. Algorithms. 2023 Feb 6;16(2):88.
6. Hankey GJ. Potential new risk factors for ischemic stroke: what is their potential?. Stroke. 2006 Aug 1;37(8):2181-8.
7. WHO MONICA Project Principal Investigators. The World Health Organization MONICA Project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. Journal of clinical epidemiology. 1988 Jan 1;41(2):105-14.
8. Vymazal J, Rulseh AM, Keller J, Janouskova L. Comparison of CT and MR imaging in ischemic stroke. Insights into imaging. 2012 Dec;3(6):619-27.
9. Saxena K, Sharma R. Efficient heart disease prediction system. Procedia Computer Science. 2016 Jan 1;85:962-9.
10. Saw M, Saxena T, Kaithwas S, Yadav R, Lal N. Estimation of prediction for getting heart disease using logistic regression model of machine learning. In2020 International Conference on Computer Communication and In formatics (ICCCI) 2020 Jan 22 (pp. 1-6). IEEE.
11. Alaa AM, Bolton T, Di Angelantonio E, Rudd JH, Van der Schaar M. Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants. PloS one. 2019 May 15;14(5):e0213653.
12. Abd Mizwar AR, Sunyoto A, Arief MR. Stroke Prediction using Machine Learning Method with Extreme Gradient Boosting Algorithm. MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer. 2022 Jul 23;21(3):595-606.
13. Uddin S, Haque I, Lu H, Moni MA, Gide E. Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. Scientific Reports. 2022 Apr 15;12(1):1-1.
14. Deng Z, Zhu X, Cheng D, Zong M, Zhang S. Efficient kNN classification algorithm for big data. Neurocomputing. 2016 Jun 26;195:143-8.
15. Zhao D, Hu X, Xiong S, Tian J, Xiang J, Zhou J, Li H. K-means clustering and kNN classification based on negative databases. Applied soft computing. 2021 Oct 1;110:107732.
16. Panda NR. A Review on Logistic Regression in Medical Research. National Journal of Community Medicine. 2022 Apr 30;13(04):265-70.
17. Cheng W, Hüllermeier E. Combining instance-based learning and logistic regression for multilabel classification. Machine Learning. 2009 Sep;76:211-25.
18. Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, Liu T, Wang X, Wang G, Cai J, Chen T. Recent advances in convolutional neural networks. Pattern recognition. 2018 May 1;77:354-77.
19. Aghdam HH, Heravi EJ. Guide to convolutional neural networks. New York, NY: Springer. 2017;10(978-973):51.
20. Zhou J, Cui G, Hu S, Zhang Z, Yang C, Liu Z, Wang L, Li C, Sun M. Graph neural networks: A review of methods and applications. AI open. 2020 Jan 1;1:57-81.
21. Qadir Z, Ever E, Batunlu C. Use of neural network based prediction algorithms for powering up smart portable accessories. Neural Processing Letters. 2021 Feb;53:721-56.
22. Baldi P, Brunak S, Chauvin Y, Andersen CA, Nielsen H. Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics. 2000 May 1;16(5):412-24.
23. Pathak S, Mishra I, Swetapadma A. An assessment of decision tree based classification and regression algorithms. In2018 3rd International Conference on Inventive Computation Technologies (ICICT) 2018 Nov 15 (pp. 92-95). IEEE.
24. Song YY, Ying LU. Decision tree methods: applications for classification and prediction. Shanghai archives of psychiatry. 2015 Apr 4;27(2):130.

25. Priyam A, Abhijeeta GR, Rathee A, Srivastava S. Comparative analysis of decision tree classification algorithms. International Journal of current engineering and technology. 2013 Jun 2;3(2):334-7.

26. Charbuty B, Abdulazeez A. Classification based on decision tree algorithm for machine learning. Journal of Applied Science and Technology Trends. 2021 Mar 24;2(01):20-8.

27. Vanisree K, Singaraju J. Decision support system for congenital heart disease diagnosis based on signs and symptoms using neural networks. International Journal of computer applications. 2011 Apr 6;19(6):6-12.

28. Absar N, Das EK, Shoma SN, Khandaker MU, Miraz MH, Faruque MR, Tamam N, Sulieman A, Pathan RK. The efficacy of machine-learning-supported smart system for heart disease prediction. InHealthcare 2022 Jun 18 (Vol. 10, No. 6, p. 1137). MDPI.

29. Drożdż K, Nabrdalik K, Kwiendacz H, Hendel M, Olejarz A, Tomasik A, Bartman W, Nalepa J, Gumprecht J, Lip GY. Risk factors for cardiovascular disease in patients with metabolic-associated fatty liver disease: a machine learning approach. Cardiovascular Diabetology. 2022 Dec;21(1):1-2.

30. Rajathi S, Radhamani G. Prediction and analysis of Rheumatic heart disease using kNN classification with ACO. In2016 International Conference on Data Mining and Advanced Computing (SAPIENCE) 2016 Mar 16 (pp. 68-73). IEEE.

31. Shah D, Patel S, Bharti SK. Heart disease prediction using machine learning techniques. SN Computer Science. 2020 Nov;1:1-6.

32. Alotaibi FS. Implementation of machine learning model to predict heart failure disease. International Journal of Advanced Computer Science and Applications. 2019;10(6).