International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Predicting Stroke Risk: An Effective Stroke Prediction Model Based on Neural Networks

Dhivya.N[1], Shalini.M [2]

[1](MCA, M.Phil (Ph.D))Assistant Professor, Department of MCA, Vivekananda institute of information and management studies.
[2]PG Student, Department of MCA, Vivekananda institute of information and management studies.

ABSTRACT:

This research uses machine learning to develop and evaluate models for long-term risk prediction of stroke occurrence. The main contribution is a stacking method that achieves high performance and accuracy. Four models are trained using Random Forest Classification and Voting Classifier, with Random Forest being the best performing algorithm with an accuracy of approximately percent. The dataset used in the development of the method is the open-access Stroke Prediction dataset. Early recognition of stroke symptoms can significantly contribute to stroke prediction and promote a healthy life.

1. INTRODUCTION

A stroke is defined as an acute neurological disorder of the blood vessels in the brain that occurs when the blood supply to an area of the brain stops and the brain cells are deprived of the necessary oxygen. Stroke is divided into ischemic and hemorrhagic. It can be mild to very severe with permanent or temporary damage. Haemorrhages are rare and involve the rupture of a blood vessel resulting in cerebral haemorrhage. Ischemic strokes, which are the most Common, involve the cessation of blood flow to an area of the brain due to a narrowing or blockage of an artery Factors that increase the chance of having a stroke are the existence of a similar stroke in the past, the existence of a transient stroke, the presence of myocardial infarction, and other heart diseases, such as heart failure, atria fibrillation, age high blood cholesterol, diabetes, obesity, sedentary lifestyle, alcohol consumption, blood clotting disorders, estrogens therapy and the use of euphoric substances such as cocaine and amphetamines. Stroke progresses rapidly, and its symptoms can vary. Symptoms can sometimes develop slowly and sometimes it can develop quickly. It is even possible for someone to wake up while sleeping with symptoms.

A stroke occurs with the sudden onset of one or more symptoms. The main ones are paralysis of the arms or legs (usually on one side of the body), numbness in the arms or legs or sometimes on the face, difficulty speaking, difficulty walking, dizziness, decreased vision, headache and vomiting and a drop in the angle of the mouth (crooked mouth). Finally, in severe strokes, the patient loses consciousness and falls into a coma, once the patient has had a stroke, a computerized tomography (CT) scan immediately provides a diagnosis. In the case of ischemic stroke, magnetic resonance imaging (MRI) is efficient. Other ancillary diagnostic tests are carotid triplex and cardiac triplex.

Strokes can be severe (extensive) or mild. The mortality rate and the number of affected people by this disease are expected to grow with the population of the world. But this mortality rate can prevent by early treatment and early prediction. There are some tools like the Cox proportional hazard model that are used to predict stroke. But it is unable to effectively prognosis to stroke with high dimensional data because of being a traditional method. In that case, machine learning can play a vital role in predicting stroke effectively and efficiently with a lower cost. Although some patients recover after a stroke, the vast majority continue to have problems depending on the severity of the stroke, such as memory, concentration and attention problems, difficulty speaking or understanding speech, emotional problems such as depression, loss of balance or the ability to walk, loss of sensation on one side of the body.

2. LITERATURE REVIEW AND MOTIVATION

Researchers have used machine learning techniques to predict stroke risk and severity using electronic health records (EHRs).

They have used an RF algorithm to predict stroke occurrence, achieving promising results. EHR-based prediction models provide real-time and comprehensive patient information, enabling early identification of individuals at risk.

However, few studies have conducted performance analyses of different machine learning algorithms for stroke prediction.

Heo et al. developed an NN-based model to predict ischemic stroke severity using MRI data, showing promising accuracy.

Wang et al. reviewed classification algorithms for stroke prediction, and Karthik et al. used bootstrap resampling to generate multiple samples.

Researchers used Electronic Health Records (EHRs) and various algorithms to improve model accuracy and reliability.

The findings demonstrate machine learning's potential in stroke management and patient outcomes.

3. EXPERIMENTAL SETUP

This section describes the experimental setup used in this work and highlights the background to the experiments.

3.1. DATASET

This work uses the Healthcare-dataset-stroke-data.csv20 dataset, which contains information about stroke occurrences and health-related factors. The dataset is used in data analysis and machine learning tasks to study and predict stroke risks.

The dataset can analyze the relationships between factors and stroke occurrences and develop predictive models for identifying individuals at higher risk. Before use, the dataset is pre-processed and cleaned, handling missing values, normalizing numerical features, and encoding categorical variables. Exploratory data analysis techniques can also be applied to gain insights into data distribution and potential patterns or correlations.

3.2. PROPOSED MODEL

The proposed experimental approach is illustrated in a flowchart, outlining the steps and machine learning models.

4. RESULTS AND DISCUSSION

The study analyzed stroke prediction using various models, with accuracies ranging from NB to NN. The LR, SVM, KNN, RF, and NN algorithms showed comparable high accuracies, outperforming NB, DT, and XGBoost.

LR, SVM, KNN, RF, and NN were found to be effective in predicting stroke risk. LR handles linear relationships effectively, SVM handles both linear and nonlinear relationships, KNN captures local patterns, RF combines multiple DTs, and NN excels in learning hierarchical representations of input data.

The NB and DT models, while achieving slightly lower accuracy levels, offer viable options for stroke prediction due to their ability to learn relevant features from data, but may suffer from over fitting and limited performance with complex relationships, interpretability, computational requirements, and the specific goals of the stroke prediction task.

The metrics used to carry out performance analysis of the algorithm are the accuracy score, precision (P), recall (R), F-measure,

and receiver operating characteristic (ROC) curve. Graphical comparison of the various algorithms using these metrics is presented

in Figs. 2e4. Model Validation: It is essential to conduct thorough model validation with separate datasets to guarantee robust performance and generalizability across various patient groups. In this research, the training set was split into k subsets, or folds, for k-fold cross validation. The model was trained k ¼ 5 times, using k_1 folds for training and 1-fold for validation each time.

This process is repeated for every fold, and the performance indicators are averaged over all folds. A split ratio of 80%/20% for the training and test sets was used in this research. The L2 regularization method was applied to avoid over fitting. Overly complex models are penalized by regularization, and early stopping ends the training process when the model performance begins to deteriorate on the validation set.

5. COMPARATIVE ANALYSIS OF TECHNIQUES

The study analyzed various stroke prediction techniques, including Naive Bayes (NB), Logistic Regression (LR), Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Decision Tree (DT), and Random Forest (RF). NB achieved an accuracy of 88.12%, while LR achieved 95.04%. SVM and KNN showed similar accuracy, while Decision Tree achieved 91.15% and 95.17% respectively. RF showed competitive performance in capturing complex patterns associated with stroke risk, despite its lower accuracy compared to other techniques.

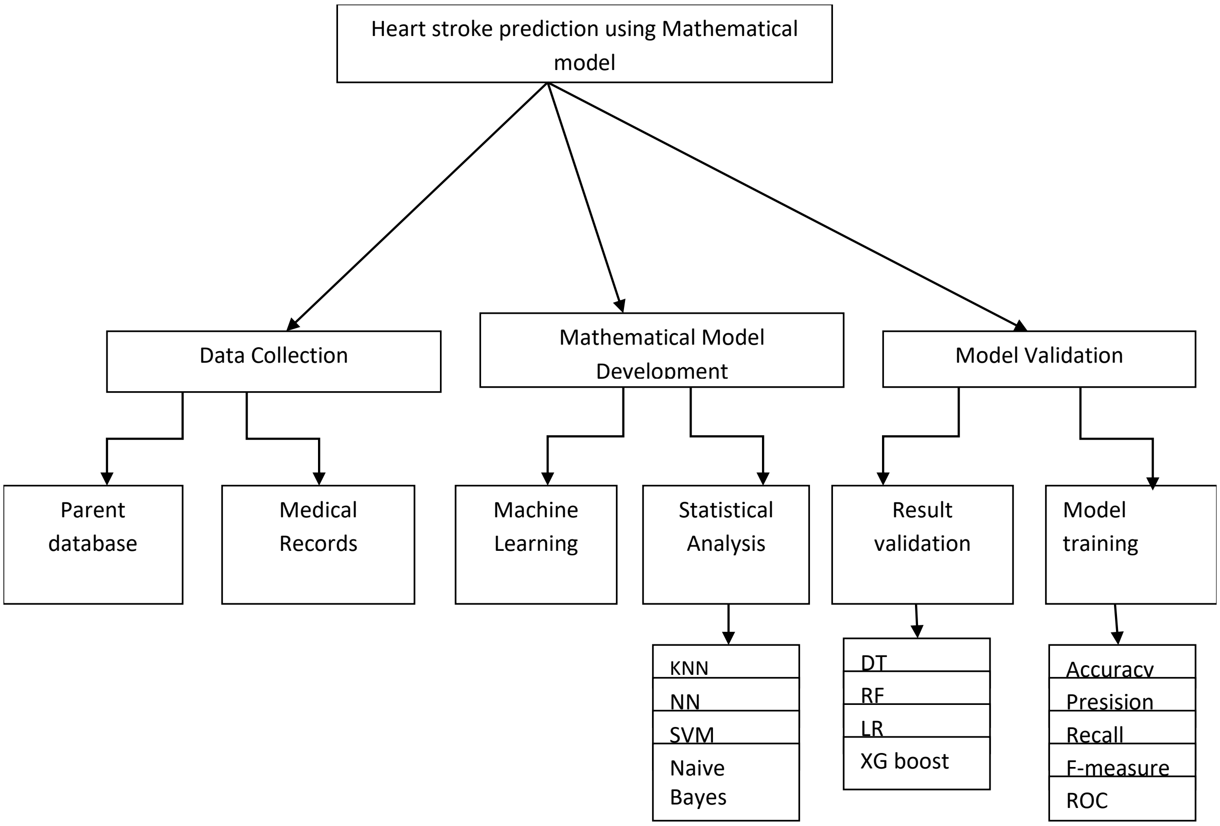
The study found that XGBoost, a stroke prediction model, achieved a high accuracy of 94.65%, while the neural network model, which achieved a 95.45% accuracy, demonstrated the highest efficacy in stroke risk assessment, suggesting its potential for personalized preventive strategies.

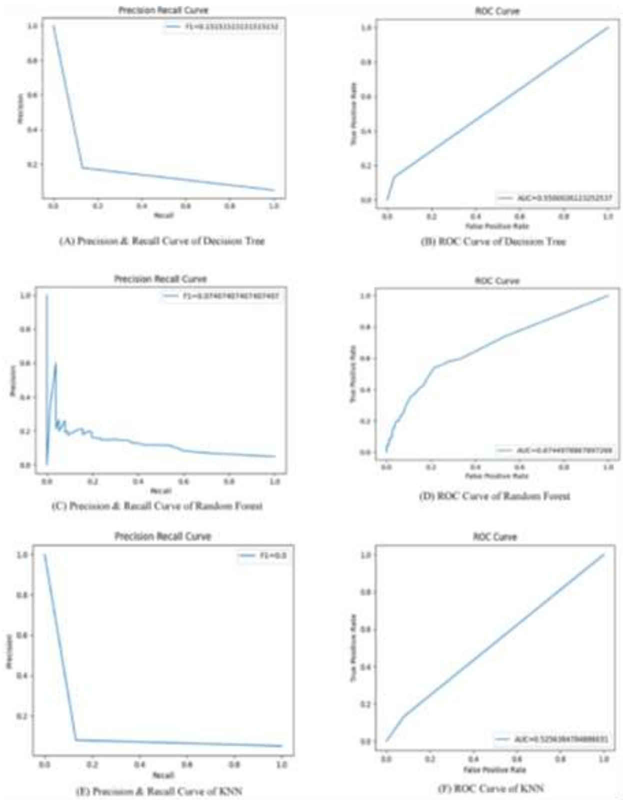
6. CONCLUSIONS

This section presents the conclusion of this work by highlighting the challenges, limitations and future scope of the presented work.

6.1. CHALLENGES AND LIMITATIONS OF THE RESEARCH CONDUCTED IN THIS STUDY

The study highlights the challenges of using machine learning algorithms to predict heart stroke risk, including missing clinical information, limited dataset size, and the need for larger real datasets. It suggests using ensemble models, enhancing data quality, and applying the SMOTE algorithm for imbalanced datasets.





(A) Precision & Recall Curve of Decision Tree
(B) ROC Curve of Decision Tree
(C) Precision & Recall Curve of Random Forest
(D) ROC Curve of Random Forest
(E) Precision & Recall Curve of KNN
(F) ROC Curve of KNN

6.2. CONCLUSION AND FUTURE SCOPE

The research aimed to evaluate the predictive performance of various machine learning algorithms on a heart stroke dataset through rigorous experimentation and evaluation.

The NN model was found to be the most effective technique for accurately predicting stroke occurrences, with NB showing the lowest accuracy.

This highlights the power of NNs in early detection and prevention strategies. Future research should explore optimizing NNs for medical research and healthcare. Future work could include integrating advanced pre-processing techniques, feature engineering methodologies, and incorporating more diverse datasets. Understanding the interpretability of neural networks could also help improve stroke prediction models.

AUTHOR CONTRIBUTIONS

All authors contributed to the manuscript, with AG and NM interpreting ideas, NJ, SM, FA, KAG, SNM analyzing results, and NM applying concepts to the proposed model.

REFERENCES

1. AlSharabi K, Bin Salamah Y, Abdurraqeeb AM, et al. EEG signal processing for Alzheimer's disorders using discrete wavelet transform and machine learning approaches. IEEE Access. 2022;10:89781e89797. https://doi.org/10.1109ACCESS.2022.3198988.

2. Fang G, Xu P, Liu WB. Automated ischemic stroke subtyping based on machine learning approach. IEEE Access. 2020;8:118426e118432. https://doi.org/10.1109/ACCESS.2020.3004977.

3. Kang K, Park TH, Kim N, et al. Recurrent stroke, myocardial infarction, and major vascular events during the first year after acute ischemic stroke: the multicenter prospective observational study about recurrence and its determinants after acute ischemic stroke I. J Stroke Cerebrovasc Dis. 2016;25(3): 656e664. https://doi.org/10.1016/j.jstrokecerebrovasdis.2015.11.036.

4. Zhang ZH, Beck MW, Winkler DA, et al. Opening the black box of neural networks: methods for interpreting neural network models in clinical applications. Ann Transl Med. 2018;6(11):216. https://doi.org/10.21037/atm.2018.05.32.

5. Han TY, Nebelung S, Pedersoli F, et al. Advancing diagnostic performance and clinical usability of neural networks via adversarial training and dual batch normalization. Nat Commun. 2021;12(1):4315. https://doi.org/10.1038/s41467-021-24464-3.

6. Yang YJ, Zheng J, Du ZZ, et al. Accurate prediction of stroke for hypertensive patients based on medical big data and machine learning algorithms: retrospective study. JMIR Med Inform. 2021;9(11):e30277. https://doi.org/10.2196/30277.

7. Zhu EZ, Chen ZH, Ai P, et al. Analyzing and predicting the risk of death in stroke patients using machine learning. Front Neurol. 2023;14:1096153. https://doi.org/10.3389/fneur.2023.1096153.

8. Mia R, Khanam S, Mahjabeen A, et al. Exploring machine learning for predictingcerebral stroke: a study in discovery. Electronics. 2024;13(4):686. https://doi.org/10.3390/electronics13040686.

9. Dritsas E, Trigka M. Stroke risk prediction with machine learning techniques.Sensors. 2022;22(13):4670. https://doi.org/10.3390/s22134670.

10. Mridha K, Ghimire S, Shin J, et al. Automated stroke prediction using machinelearning: an explainable and exploratory study with a web application for earlyintervention. IEEE Access. 2023;11:52288e52308. https://doi.org/10.1109/ACCESS.2023.3278273.

11. Heo J, Yoon JG, Park H, et al. Machine learning-based model for prediction of outcomes in acute stroke. Stroke. 2019;50(5):1263e1265. https://doi.org/10.1161/STROKEAHA.118.024293.

12. Wang WJ, Kiik M, Peek N, et al. A systematic review of machine learning models for predicting outcomes of stroke with structured data. PLoS One. 2020;15(6):e0234722. https://doi.org/10.1371/journal.pone.0234722.

13. Karthick K, Aruna SK, Manikandan R. Development and evaluation of the bootstrap resampling technique based statistical prediction model for Covid-19 real time data: a data driven approach. J Interdiscipl Math. 2022;25(3):615e627. https://doi.org/10.1080/09720502.2021.2012890.

14. Chantamit-O-Pas P, Goyal M. Prediction of stroke using deep learning model. In: Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14e18, 2017, Proceedings, Part V 24. Berlin, Germany: Springer International Publishing; 2017:774e781.

15. Conference welcome map. In: 2016 International Conference On Control, Instrumentation, Communication And Computational Technologies (ICCICCT). India: Kumaracoil; 2016:1. https://doi.org/10.1109/ICCICCT.2016.7988069, 1.

16. Al-Shammari NK, Alzamil AA, Albadarn M, et al. Cardiac stroke prediction framework using hybrid optimization algorithm under DNN. Eng Technol Appl Sci Res. 2021;11(4):7436e7441. https://doi.org/10.48084/etasr.4277.

17. Emon MU, Keya MS, Meghla TI, et al. Performance analysis of machine learning approaches in stroke prediction. In: 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA). 2020: 1464e1469. https://doi.org/10.1109/ICECA49313.2020.9297525. Coimbatore, India.

18. Tazin T, Alam MN, Dola NN, et al. Stroke disease detection and prediction using robust learning approaches. J Healthc Eng. 2021;2021:7633381. https://doi.org/ 10.1155/2021/7633381.

19. Dev S, Wang HW, Nwosu CS, et al. A predictive analytics approach for stroke prediction using machine learning and neural networks. Healthc Anal. 2022;2:

20. Fedesoriano. Stroke Prediction Dataset. Available