

Artificial Intelligence-Based Brain Stroke Prediction: A Review of Machine Learning, Deep Learning, and Emerging Healthcare Technologies

Kuldeep hariyale¹, Aditya Patel²

^{1,2}Sanjeev Agrawal Global Educational (SAGE) University Bhopal

kuldeephariyale5818@gmail.com

Abstract

Brain strokes represent major global health crises causing mortality plus long term disability. Early prediction plus rapid diagnostic identification assist doctors during treatment reducing fatality rates plus improving outcomes for human patients. Recent technical progress inside AI plus Machine Learning plus Deep Learning improved calculation speed plus overall precision for medical prediction software programs. This text synthesizes studies addressing stroke identification trends covering data years 2020 through 2025. Several logic algorithms including Logistic Regression plus Vector Machines plus Decision Trees plus Random Forests plus blended model systems are compared herein. Besides deep architecture pathways like Convolutional Neural Networks plus EfficientNet plus merged hybrid layouts focused on MRI plus CT scan visual data analysis appear within these pages. Our overview looks toward new sectors like Explainable AI plus Federated Learning plus Blockchain plus Digital Twins plus Generative Artificial Intelligence structures targeting smarter medical platforms. Persistent hurdles regarding data validation plus category unevenness plus opaque model functions plus security protection plus real world hospital implementation receive objective analysis. Ultimately future research targets developing reliable plus open plus customized stroke predictive architectures. Findings suggest linking varied medical datasets inside open security focused software frameworks promotes higher effectiveness during future health predictive endeavors worldwide.

Keywords: Brain Stroke Prediction, Machine Learning, Deep Learning, Medical Imaging, Explainable AI, Federated Learning, Healthcare Analytics.

I. Introduction

Stroke creates severe neurological harm because blood supply stops reaching human brain tissue samples. Academic medical studies identify severe stroke events among primary causes regarding global sudden deaths plus ongoing permanent physical disability. Spotting dangerous health risk markers quite early lowers patient fatality odds while bolsters positive healing medical therapeutic treatment output success. Traditional diagnostic clinical office work involves inspecting symptoms through complex neuroimaging gear; multiple diagnostic methods often prove slow since high professional specialized analytical expertise feels strictly necessary. Intelligent computing automated software offers improved high quality support towards current clinical analytical efforts, permitting accurate prediction plus precise medical diagnosis via thorough diagnostic data information pattern interpretation. Advanced neural machine processing models detect vital strong biological links inside metrics including advanced age, vascular hypertension, glucose diabetes, daily nicotine use, high weight ratio, plus serious cardiovascular conditions. Numerous technical review findings report forest modeling logic plus deep system networks perform strong during routine health stroke assessment events. This article describes methodology comparison, identifies logic gaps.

II. Literature Review

Most recent studies on stroke prediction have been using machine learning approaches. In the early stages of research a variety of classifiers have been used, including simple ones like Logistic Regression, Decision Trees, K-NN, Support Vector Machines and Random Forests. Among those the performance of Random Forest and Support Vector Machine have been outstanding due to their ability to handle non-linear relationships in complex data such as healthcare. Recently, in a number of systematic reviews, Random Forest has been reported as one of the most frequently cited high-performing models for stroke prediction.

Several studies have introduced ensemble learning to improve the prediction accuracy of stroke-related diseases. These approaches are categorized into boosting, bagging, stacking, and soft-voting classifiers. In general, these models

combine multiple classifiers and can outperform their individual counterparts by decreasing variance and improving the generalization ability of a model.

Stroke prediction also benefited from the advent of deep learning. In medical applications including stroke prediction, the use of convolutional neural networks (CNNs) for the analysis of medical images, such as MRI and CT scans, is gaining increasing attention. The fact that CNNs are able to automatically extract hierarchical features from input data without the need for manual feature design makes them particularly suitable for the analysis of medical images. Moreover, a wide range of architectures have been proposed, including EfficientNet and hybrid models, that have been designed to increase the accuracy of lesion detection, tissue classification and even the assessment of the severity of strokes. According to several recent reviews, the use of deep learning, including classification, segmentation and object detection, for the diagnosis of strokes is currently an area of growing research.

In addition to the classification accuracy, increasing attention is being paid to the explanations that a model is able to provide as well as to the means by which a model can be used to protect patient data. There are a number of Explainable AI (XAI) techniques such as SHAP and LIME that have been designed to increase transparency of deep learning models. Furthermore, techniques such as Federated Learning and the use of Blockchain in healthcare analytics are capable of providing secure and privacy-preserving healthcare analytics. Such techniques are expected to play a major role in the development of future intelligent healthcare ecosystems.

Ref.	Author & Year	Technique Used	Dataset Type	Key Findings	Limitations	Research Insight
1	Emon et al. (2020)	ML Classifiers (RF, SVM, DT)	Clinical Data	Compared multiple ML models for stroke prediction	Limited dataset size	Ensemble models outperform individual classifiers
2	Vamsi et al. (2020)	Machine Learning	Clinical Data	Stroke severity prediction achieved promising results	Lack of explainability	Severity prediction is an emerging research area
3	Sailasya & Kumari (2021)	Classification Algorithms	Clinical Data	Random Forest achieved superior performance	Class imbalance issue	Feature selection significantly impacts accuracy
4	Krishna et al. (2021)	ML Models	Clinical Data	Early stroke detection using supervised learning	Small sample size	Early diagnosis can reduce mortality risk
5	Tazin et al. (2021)	Robust Learning Models	Clinical Data	Improved prediction performance through robust learning	Generalization issue	Data quality strongly influences outcomes
6	Akter et al. (2022)	Machine Learning	Clinical Data	Effective stroke disease detection framework	Limited interpretability	ML remains useful for structured healthcare data
7	Devaki & Rao (2022)	Ensemble Framework	Clinical Data	Improved prediction accuracy using ensemble learning	Increased computational complexity	Ensemble learning improves robustness

8	Dritsas & Trigka (2022)	ML Techniques	Clinical Data	Comprehensive risk prediction analysis	Dataset dependency	Random Forest consistently performs well
9	Madhavi et al. (2022)	CNN	MRI Images	Deep learning achieved high detection accuracy	Large training data requirement	CNNs are effective for imaging applications
10	Mostafa et al. (2022)	Ensemble Classifier	Clinical Data	Ensemble approach improved classification performance	Reduced transparency	Hybrid ensembles offer better prediction stability
11	Almubark (2023)	Machine Learning	Clinical Data	ML models successfully predicted stroke risk	Lack of multimodal analysis	Clinical features remain strong predictors
12	Bathla & Kumar (2023)	Hybrid ML + Feature Selection	Clinical Data	Combined feature selection improved accuracy	Limited external validation	Feature optimization improves model efficiency
13	Bhalla et al. (2023)	Comparative ML Analysis	Clinical Data	Compared multiple classifiers	Dataset-specific findings	No single classifier is universally optimal
14	Kumari & Garg (2023)	Explainable ML	Clinical Data	Investigated model interpretability	Limited XAI techniques	Explainable AI is crucial for healthcare adoption
15	Rehman (2023)	Deep Learning + ADASYN	Clinical Data	Improved minority-class prediction	Computational overhead	Class imbalance handling is essential

III. Comparative Analysis of Existing Approaches

Table I summarizes the major AI techniques used in stroke prediction.

Table I. Comparison of Stroke Prediction Approaches

Approach	Advantages	Limitations
Logistic Regression	Simple, interpretable	Limited nonlinear modeling
SVM	Good generalization	Computationally expensive
Random Forest	High accuracy, robust	Reduced interpretability
Ensemble Learning	Improved performance	Increased complexity
CNN	Excellent image analysis	Large data requirement
Vision Transformer	Strong feature learning	High computational cost

Traditional machine learning models are suitable for structured clinical datasets, whereas deep learning approaches are particularly effective for medical imaging. Ensemble models generally achieve higher predictive performance, but their complexity may limit clinical adoption. Therefore, balancing accuracy and interpretability remains a key challenge.

IV. Research Challenges

Despite significant progress, several challenges continue to hinder the practical deployment of AI-based stroke prediction systems.

The first challenge involves data availability and quality. Most studies utilize relatively small datasets that may not represent diverse patient populations. Missing values, noise, and inconsistent records further affect model reliability.

Class imbalance is another major issue because stroke-positive cases are often significantly fewer than non-stroke cases. This imbalance may cause predictive models to favor majority classes and increase false-negative predictions.

Model interpretability also remains a critical concern. Although deep learning models achieve high accuracy, they often operate as black-box systems. Clinicians require transparent explanations before relying on AI-assisted decisions.

Privacy and security concerns represent additional barriers. Healthcare data are highly sensitive, making secure data sharing essential. Technologies such as Federated Learning and Blockchain have been proposed to address these concerns; however, scalability and implementation challenges remain unresolved.

V. Future Research Directions

Future research should focus on the development of multimodal learning systems capable of integrating clinical records, laboratory reports, MRI scans, CT images, wearable sensor data, and genomic information. Such systems can provide a more comprehensive representation of patient health and improve prediction accuracy.

Explainable Artificial Intelligence will play an increasingly important role in healthcare applications. Future models should provide transparent reasoning processes that support clinician trust and regulatory compliance.

Federated Learning, Blockchain, and Differential Privacy are expected to facilitate secure collaboration among healthcare institutions while preserving patient confidentiality. Additionally, Digital Twin technology and Generative AI may enable personalized healthcare, real-time monitoring, and predictive disease management.

The integration of these technologies will contribute to the development of Healthcare 5.0 ecosystems characterized by intelligent, secure, and patient-centered decision support systems.

VI. Conclusion

This review examined recent advances in AI-based brain stroke prediction, focusing on machine learning, deep learning, medical imaging analytics, explainable AI, and privacy-preserving healthcare technologies. The literature demonstrates that AI techniques can significantly improve early stroke detection and clinical decision-making. Traditional machine learning algorithms remain effective for structured healthcare data, whereas deep learning architectures provide superior performance in medical imaging applications.

However, challenges related to dataset quality, class imbalance, interpretability, privacy, and clinical validation continue to limit widespread deployment. Future research should emphasize multimodal learning, explainable AI, federated healthcare systems, and personalized medicine. The integration of these emerging technologies has the potential to transform stroke prediction from a reactive diagnostic process into a proactive and preventive healthcare framework.

References

- [1] M. U. Emon, M. S. Keya, T. I. Meghla, Md. M. Rahman, M. S. Al Mamun, and M. S. Kaiser, "Performance Analysis of Machine Learning Approaches in Stroke Prediction," in *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2020, pp. 1464–1469, doi:10.1109/ICECA49313.2020.9297525.
- [2] B. Vamsi, D. Bhattacharyya, and D. Midhunchakkavarthy, "Prediction of Brain Stroke Severity Using Machine Learning," *Revue d'Intelligence Artificielle*, vol. 34, no. 6, pp. 753–760, 2020, doi:10.18280/RIA.340609.
- [3] G. Sailasya and G. L. A. Kumari, "Analyzing the Performance of Stroke Prediction using ML Classification Algorithms," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, 2021, doi:10.14569/IJACSA.2021.0120662.
- [4] V. G. Krishna, J. S. Kiran, D. P. Rao, G. C. Babu, and G. J. Babu, "Early Detection of Brain Stroke using Machine Learning Techniques," in *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)*, 2021, doi:10.1109/ICOSEC51865.2021.9591840.
- [5] T. Tazin, M. N. Alam, N. N. Dola, M. S. Bari, S. Bourouis, and M. M. Khan, "Stroke Disease Detection and Prediction Using Robust Learning Approaches," *Journal of Healthcare Engineering*, vol. 2021, Article ID 7633381, 2021, doi:10.1155/2021/7633381.
- [6] B. Akter, A. Rajbongshi, S. Sazzad, R. Shakil, J. Biswas, and U. Sara, "A Machine Learning Approach to Detect the Brain Stroke Disease," in *2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2022, doi:10.1109/ICSSIT53264.2022.9716345.
- [7] A. M. Devaki and C. V. G. Rao, "An Ensemble Framework for Improving Brain Stroke Prediction Performance," in *International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, 2022, doi:10.1109/ICEEICT53079.2022.9768579.
- [8] E. Dritsas and M. Trigka, "Stroke Risk Prediction with Machine Learning Techniques," *Sensors*, vol. 22, no. 13, p. 4670, 2022, doi:10.3390/S22134670.
- [9] K. R. Madhavi, K. Kovuri, J. Avanija, M. Sakthivel, and S. Kaleru, "Brain Stroke Prediction Using Deep Learning: A CNN Approach," in *2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2022, doi:10.1109/ICIRCA54612.2022.9985596.
- [10] S. A. Mostafa, D. S. Elzanfaly, and A. Yakoub, "A Machine Learning Ensemble Classifier for Prediction of Brain Strokes," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 12, 2022, doi:10.14569/IJACSA.2022.0131232.
- [11] I. Almubark, "Brain Stroke Prediction Using Machine Learning Techniques," in *2023 IEEE International Conference on Big Data (BigData Congress)*, 2023, doi:10.1109/BIGDATA59044.2023.10386474.
- [12] P. Bathla and R. Kumar, "A Hybrid System to Predict Brain Stroke Using a Combined Feature Selection and Classifier," *Intelligent Medicine*, vol. 3, no. 3, pp. 162–170, 2023, doi:10.1016/J.IMED.2023.06.002.
- [13] S. Bhalla, A. Kumar, and R. Kushwaha, "Comparative Analysis on Brain Stroke Prediction using Machine Learning," in *International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2023, doi:10.1109/ICCCNT56998.2023.10308218.
- [14] R. Kumari and H. Garg, "Interpretation and Analysis of Machine Learning Models for Brain Stroke Prediction," in *2023 6th International Conference on Information Systems and Computer Networks (ISCON)*, 2023, doi:10.1109/ISCON57294.2023.10112188.
- [15] A. Rehman, "Brain Stroke Prediction through Deep Learning Techniques with ADASYN Strategy," in *International Conference on Developments in eSystems Engineering (DeSE)*, 2023, doi:10.1109/DESE60595.2023.10469013.