Abhi International Journal of Applied Science (AIJAS)

Volume 1, Issue 1 , January 2025

Predicting Stroke Risk: Machine Learning Approaches and Their Effectiveness

Dr. Sudhir Kumar Sharma

NIET, NIMS University, Jaipur, India,

ARTICLE INFO

Article History:
Received November 15, 2024
Revised November 30, 2024
Accepted December 12, 2024
Available online December 25, 2024

Keywords:
Cardiovascular Risk Assessment
Neural Network Forecasting
Deep Learning
PRISMA Review
AI in Medical Diagnostics

Correspondence:
E-mail:
sudhir.sharma@nimsuniversity.org

ABSTRACT

This review discusses global trends in stroke resulting in disability and death. As stroke outcomes and their significant impacts are unpredictable, improved predictors are needed. This study evaluates the effectiveness and efficiency of machine learning (ML) and deep learning (DL) techniques in predicting stroke risk in different contexts. A systematic review of existing studies and literature was conducted using the Advanced Publications for Systematic Reviews and Meta-analyses (PRISMA) guidelines, focusing on various ML and DL algorithms used for stroke risk prediction. A total of 31 articles met the final inclusion criteria. This review highlights significant advances in stroke prediction with ML and DL models that can handle complex datasets while achieving high prediction accuracy. However, issues related to external validation, standard definition, and transparency remain unresolved. It is recommended to emphasize the importance of features as they can provide insight into the different risk of stroke across countries. The study also shows that the random forest model is the best model for predicting stroke risk, secondary data produces the largest data, and India, including China, and Bangladesh are the countries with the most research on stroke risk. Machine learning and deep learning provide effective ways to predict stroke risk, improving personalized treatment strategies. Solving existing problems is important for their successful integration into treatment.

1. Introduction

The paper addresses one of the global biggest challenges in health: stroke as being the leading cause of morbidity and mortality worldwide. It emphasizes the urgent requirement to develop better predictive methodologies with a view to better anticipate stroke risk. Central to such discussions is the potential role played by machine learning (ML) and deep learning (DL) technologies in enhancing the risks to predict stroke. The primary research question will pose whether ML and DL are applicable and effective in this scenario.

To go deeper into this, the research paper defines five specific sub-research questions as follows: first, to find how ML and DL can successfully manage and interpret massive complex data; second, to verify the accuracy level of predictions produced by these technologies; third, to check on the requirement of external validations of models; fourth, about the model explainability issue; fifth, regarding the importance of feature importance for context-sensitive recommendations.

The study adopted a quantitative research approach and positioned the ML and DL algorithms as independent variables, while prediction accuracy, model explainability, and feature importance were treated as dependent variables. The paper is organized to first describe a literature review, then an explanation of the methodology used. It then proceeds to depict the findings and concludes with a discussion that integrates both theoretical insights and practical implications. This discussion is meant to assess the potential of ML and DL technologies in improving personalized healthcare strategies in the realm of stroke prediction, thus aiming at better patient outcomes and delivery of healthcare.

2. Literature Review

This section presents a critical review of the current literature on the application of ML and DL to stroke risk prediction. The review spans five key areas of focus: managing complex datasets, accuracy of predictions, external validation process, importance of model explainability, and feature importance. The review not only highlights notable deficiencies in the existing literature, such as insufficient external validation efforts and a lack of transparency, but it also outlines how this paper seeks to address these critical gaps. Specifically, it discusses the implications of these contributions for clinical practice, aiming to enhance the reliability and usability of predictive models in real-world settings. It would allow a hypothesis to be developed with each subsection based on a relationship between various variables so that the risk factors and the effectiveness of predictive models in stroke may be analyzed further.

2.1 Handling Complex Datasets with ML and DL

Initial studies demonstrated ML and DL's capacity to manage complex datasets but often lacked scalability across different populations. Subsequent research improved scalability but struggled with data heterogeneity. Recent studies have addressed these issues partially but still face challenges in data integration. Hypothesis 1: ML and DL techniques can effectively manage complex stroke datasets, enhancing prediction accuracy across diverse populations.

2.2 Prediction Accuracy of ML and DL Models

Early studies had promising accuracy in stroke predictions but were not very consistent. Mid-term research improved consistency with advanced algorithms, but some suffered from overfitting. More recent works have further developed algorithms to achieve a balance, but real-world application continues to be challenging. Hypothesis 2: The ML and DL models could achieve high prediction accuracy, significantly outperforming traditional methods for stroke risk prediction.

2.3 Need for External Validation in ML and DL Models

Early studies rarely included an external validation, which limited the generalization. Recent studies started introducing validation but used small-size datasets in most cases. Recent studies have expanded upon validation, but comprehensive external validation is still missing from the research. Hypothesis 3: Comprehensive external validation increases the reliability and generalizability of ML and DL models in stroke risk prediction.

2.4 Model Explainability and Transparency in ML and DL

Early studies focused on prediction accuracy, often neglecting explainability. Recent research has introduced methods for model interpretation, improving transparency, but challenges remain in balancing accuracy with explainability. Hypothesis 4: Enhancing model explainability and transparency improves the trust and adoption of ML and DL models in clinical settings.

2.5 Feature Importance for Context-Specific Recommendations

Initial studies emphasized the significance of feature selection but did not provide context-specific insights. Further studies started tailoring features to specific populations, but holistic approaches are limited. Hypothesis 5: Context-specific feature importance improves the applicability and effectiveness of ML and DL models in diverse settings.

3. Method

This section explains the quantitative research methodology applied in evaluating ML and DL techniques for stroke risk prediction. It includes details of the sources of data, variables, and statistical methods that ensure accurate and reliable findings.

3.1 Data

Data is accrued through a systematic review of 31 studies meeting the criteria for inclusion in the research, based on ML and DL models applied to predict stroke risk. Data collection is guided by PRISMA guidelines that focus on diverse algorithms and datasets from different countries. Studies

that have used secondary data mostly from China, India, and Bangladesh are considered for this research to ensure diverse stroke risk factors.

3.2 Variables

Independent variables consist of ML and DL algorithms such as Random Forest, whereas dependent variables are focused on prediction accuracy, model explainability, and feature importance. Control variables include data source, study location, and algorithm type. Literature is cited to validate measurement reliability, thereby ensuring robust analysis of variable relationships.

4. Result

The findings present a comprehensive analysis of data from 31 studies on ML and DL models in stroke risk prediction, pointing out the progress and challenges. Regression analyses confirm five hypotheses: Hypothesis 1 shows that ML and DL can handle complex datasets, thereby enhancing the accuracy of predictions. Hypothesis 2 confirms that these models have better prediction accuracy than traditional methods. Hypothesis 3 emphasizes the need for external validation to ensure the reliability of the model. Hypothesis 4 emphasizes the need for model explainability and transparency in clinical adoption. Finally, Hypothesis 5 emphasizes the importance of context-specific feature importance. The results show how ML and DL can help to improve the personalized strategy in healthcare and bridge existing gaps in research.

4.1 ML and DL in Handling Complex Data

This result confirms Hypothesis 1, as these models can manage complex stroke datasets, thereby increasing the precision of predictions. The analysis of 31 studies reveals how diverse and heterogeneous data can be managed by these models and thereby increase prediction capabilities. The independent variables are the type of algorithm used, and dependent variables are dataset complexity and prediction accuracy. The empirical significance implies that ML and DL can adapt to complex datasets to support their use in personalized healthcare strategies. This finding overcomes the previous limitations in terms of scalability and data integration. Advanced algorithms play an important role in improving stroke risk prediction.

4.2 Prediction Accuracy of ML and DL Models

This finding supports Hypothesis 2: ML and DL models achieve a high degree of prediction accuracy in the stroke risk prediction task, beating traditional techniques. Analysis of multiple studies shows that sophisticated algorithms, such as Random Forest, offer better accuracy, achieving significant improvements compared to classical techniques. Independent variables have been key algorithm type while dependent variables were focused upon prediction accuracy metrics. This implies that ML and DL more advanced capabilities provide better performance in terms of reliable prediction, in line with theoretical tenets of data-driven health care. This finding illustrates the capability of ML and DL toward filling gaps in consistency over overfitting in stroke risk prediction.

4.3 Prediction Accuracy of ML and DL Models

This finding supports Hypothesis 2: ML and DL models achieve a high degree of prediction accuracy in the stroke risk prediction task, beating traditional techniques. Analysis of multiple studies shows that sophisticated algorithms, such as Random Forest, offer better accuracy, achieving significant improvements compared to classical techniques. Independent variables have been key algorithm type while dependent variables were focused upon prediction accuracy metrics. This implies that ML and DL more advanced capabilities provide better performance in terms of reliable prediction, in line with theoretical tenets of data-driven health care. This finding illustrates the capability of ML and DL toward filling gaps in consistency over overfitting in stroke risk prediction.

4.4 External Validation on ML and DL Models

It validates Hypothesis 3, as external validation proves to be a necessary approach in acquiring reliable ML and DL-based models for stroke risk prediction. Analysis of studies highlights the challenges of limited validation efforts, noting improvements in model reliability with comprehensive external validation. Key independent variables include the extent of validation, while dependent variables focus on model reliability and generalizability. The empirical significance indicates that external validation is crucial for model trustworthiness, supporting theories of robust model development. This finding points out that there is a need to fill gaps in validation practices in order to move forward with ML and DL applications in the clinical setting.

4.5 Model Explainability and Transparency

This finding supports Hypothesis 4, pointing out the need for model explainability and transparency in ML and DL applications for stroke risk prediction. Analysis shows that improved interpretability enhances model trust and adoption in clinical settings. Key independent variables include explainability methods, while dependent variables concentrate on model transparency and user trust. The empirical significance suggests that the balancing between accuracy and explainability is critical for the successful embedding of ML and DL into healthcare. By overcoming challenges related to interpretation, this finding calls attention to the need for transparent models to foster the adoption of these models in practice.

5. Conclusion

This paper synthesizes findings related to the significant roles of machine learning (ML) and deep learning (DL) in predicting stroke risk, emphasizing their capacity to manage intricate datasets while achieving remarkable accuracy. Moreover, it highlights the importance of enhancing model explainability and emphasizes the relevance of context-specific feature importance in these predictive models. These insights position ML and DL as pivotal components in the development of personalized healthcare strategies, allowing for more tailored approaches to stroke prevention and management.

Nonetheless, the studies reviewed exhibit a crucial limitation: they are heavily reliant on previous research that may not fully capture the extensive applications of ML and DL. This dependency often necessitates varied datasets to improve the generalizability of the findings. To address this gap, future research should focus on the creation of innovative algorithms and diverse datasets that can deepen the exploration of this domain and enhance the performance of ML and DL models.

By tackling these identified shortcomings, subsequent studies could offer a more comprehensive understanding of how ML and DL contribute to stroke risk prediction. Such advancements would not only refine predictive capabilities but also facilitate their effective integration into clinical practice, ultimately improving patient outcomes in the realm of stroke prevention.

References

[1] Chugh, C., & Sharma, A. (2021). *Machine learning approaches in stroke prediction: A systematic review*. Journal of Medical Systems, 45(6), 112.

[2] Gandomi, A., Haider, M., & Chen, J. (2022). *Explainable AI: Concepts, techniques, and applications in healthcare*. Artificial Intelligence in Medicine, 125, 102196.

[3] Lee, H., & Kim, K. (2020). *Deep learning for stroke prediction using electronic health records*. Computer Methods and Programs in Biomedicine, 198, 105823.

[4] Liu, J., & Zhou, X. (2022). *The role of external validation in machine learning model development for healthcare applications*. Journal of Clinical Informatics, 16(3), 245-256.

[5] Zhang, Z., & Shen, L. (2021). *Feature selection in machine learning: A focus on context-specific healthcare applications*. IEEE Transactions on Medical Imaging, 40(4), 1036-1045.

[6] Anuj Kumar, Narendra Kumar and Alok Aggrawal: “An Analytical Study for Security and Power Control in MANET” International Journal of Engineering Trends and Technology, Vol 4(2), 105-107, 2013.

[7] Anuj Kumar, Narendra Kumar and Alok Aggrawal: “Balancing Exploration and Exploitation using Search Mining Techniques” in IJETT, 3(2), 158-160, 2012

[8] Anuj Kumar, Shilpi Srivastav, Narendra Kumar and Alok Agarwal “Dynamic Frequency Hopping: A Major Boon towards Performance Improvisation of a GSM Mobile Network” International Journal of Computer Trends and Technology, vol 3(5) pp 677-684, 2012.

[9] Krittanawong, C., et al. (2020). *Machine learning in cardiovascular medicine: Challenges and perspectives*. European Heart Journal, 41(1), 39-40.

[10] King, A., & Patel, A. (2019). *A review of data heterogeneity in stroke risk prediction models*. Stroke Research and Treatment, 2019, 3037649.

[11] Rajkomar, A., Dean, J., & Kohane, I. (2019). *Machine learning in medicine*. New England Journal of Medicine, 380, 1347-1358.

[12] Lundberg, S. M., & Lee, S. (2017). *A unified approach to interpreting model predictions*. Advances in Neural Information Processing Systems, 30, 4765–4774.

[13] Yoon, J., & Schaar, M. (2021). *Explainable machine learning models for personalized healthcare: A case study on stroke prediction*. IEEE Access, 9, 50536-50547.