BRAIN STROKE PREDICTION USING MACHINE LEARNING

MANUSHA M J [1], ROOPA R [2]

[1]Student Department of MCA, BIET, Davanagere.

[2]Assistant professor, Department of MCA, BIET, Davangere.

ABSTRACT

In our study, we aim to predict the likelihood of a stroke using machine learning. We used a dataset with various patient information, such as age, lifestyle, and medical history. After cleaning and preparing the data, we identified key factors that influence stroke risk. We tested several machine learning methods, including Logistic Regression, Random Forest, Support Vector Machines, and Neural Networks. We found that the Random Forest method was the most accurate, correctly predicting strokes 89% of the time. Our analysis showed that the most important predictors of stroke are hypertension, age, and heart disease. This research demonstrates that machine learning can effectively predict strokes, potentially allowing for earlier and more targeted preventive care in medical environments.

Keywords: artificial intelligence; brain stroke prediction; deep learning

I. INTRODUCTION

Stroke is a significant global health concern, ranking among the leading causes of mortality and long-term disability. According to the World Health Organization (WHO)strokes account for approximately 11% of all deaths worldwide. The consequences of a stroke can be devastating, not only to the individuals affected but also to their families and healthcare systems. Detecting and intervening at an early stage play a crucial role in enhancing results and lessening the impact of stroke.

Traditionally, stroke risk assessment has relied on clinical evaluation and heuristic-based scoring systems such as the Framingham Stroke Risk Profile. While these methods provide valuable insights, they often lack the precision and adaptability required for personalized medicine. Advances in machine learning (ML) offer a promising alternative by leveraging large datasets to identify complex patterns and relationships that may be overlooked by conventional Machine learning, a subset of artificial intelligence.

In the field of artificial intelligence, machine learning focuses on the development of algorithms that can learn from data and utilize this knowledge to make accurate predictions. In healthcare, ML has shown great potential in various applications, including disease diagnosis, patient monitoring, and personalized treatment planning. ML has the capability to examine a broad range of factors for predicting strokes, including demographic data, medical records, lifestyle choices, and clinical data, in order to produce precise risk evaluations.

This study aims to develop and evaluate machine learning models for predicting the likelihood of stroke. By utilizing a comprehensive dataset and comparing multiple ML algorithms, We aim to determine the most efficient model for forecasting strokes. Furthermore, we will examine the importance of different predictive features to provide insights into key risk factors.

The primary objectives of this research are threefold: (1) to develop a robust predictive model for stroke using machine learning techniques, (2) to compare the performance of various ML algorithms in terms of accuracy and reliability, and (3) to identify the most significant predictors of stroke risk. Through this study, we aspire to contribute to the improvement of predictive analytics in healthcare and facilitate the establishment of proactive measures for stroke prevention.

II. RELATED WORK

The study conducted in 2010 involved Khosla, A., Cao, Y., Lin, C. C.-Y., Chiu, H. K., Hu, J., and Lee, H. This paper introduces a comprehensive machine learning framework that combines different data sources and predictive algorithms to enhance stroke prediction accuracy. By integrating clinical, imaging, and genetic data, the authors utilize logistic regression, decision trees, and support vector machines to show that a multi-source approach significantly improves prediction performance[1].

Feng, R., Zhang, W., & Zhu, Y. (2018). The authors propose an ensemble learning model that includes decision trees, random forests, and gradient boosting machines for stroke risk prediction. Their study emphasizes the significance of feature selection and cross-validation techniques, demonstrating that ensemble methods outperform single classifiers in predicting stroke risk[2].

Sharma, M., Singh, G., & Sharma, A. (2019). This review paper discusses various machine learning techniques applied to stroke prediction, such as neural networks, k-nearest neighbors, and Naïve Bayes. The authors highlight the strengths and weaknesses of each method and suggest a hybrid approach to maximize prediction accuracy[3].

Saranya, S., & Thangaraj, P. (2016). The study introduces a fuzzy k-nearest neighbor (k-NN) classifier for stroke prediction, incorporating fuzzy logic to handle uncertainty in medical data. Results show that fuzzy k-NN provides better handling of ambiguous cases compared to traditional k-NN methods[4].

Hassan, A., & Mahmood, A. (2017). The paper examines the implementation of deep learning neural networks, with a specific focus on convolutional neural networks (CNNs) and recurrent neural networks (RNNs).for stroke prediction. They demonstrate that deep learning models can effectively predict strokes[5].

Gupta and Dhingra (2019) present a study titled "An Artificial Neural Network Approach for Stroke Prediction." In their research, they develop an artificial neural network model that utilizes patient demographics, medical history, and lifestyle factors to predict stroke risk. The authors find that with appropriate feature engineering, artificial neural networks can offer robust predictive performance[6].

Lee and Kim (2020) conduct a study titled "Stroke Prediction with Big Data Analytics." Their research focuses on utilizing big data analytics and stroke risk. They leverage large-scale electronic health records (EHRs) and apply techniques such as random forests, gradient boosting, and deep learning. The study highlights the advantages of using big data in improving prediction accuracy[7].

Paul and Mukherjee (2015) propose a hybrid approach in their paper titled "Hybrid Machine Learning Techniques for Stroke Prediction." They combine genetic algorithms with support vector machines (SVMs) to enhance stroke prediction. The genetic algorithm optimizes the feature selection process, while the SVM handles the classification, resulting in improved prediction performance[8].

Zhou and Li (2019) focus in their study titled "Predicting Stroke using Machine Learning and Feature Selection Techniques." They utilize techniques such as LASSO and and random forests. The results demonstrate that feature selection significantly enhances model interpretability and accuracy[9].

Shoaib and Javed (2021) investigate the use of ensemble learning methods and advanced feature engineering techniques in their research titled "Stroke Prediction using Ensemble Learning and Feature Engineering." They explore the combination of ensemble learning methods like bagging and boosting with careful feature engineering. The authors show that this approach can yield high predictive performance for stroke risk[10].

III. METHODOLOGY

The methodology for predicting brain stroke using machine learning involves several key steps:

1. **Data Collection**: Amass thorough datasets comprising population data, medical histories, daily schedules, and diagnostic findings pertinent to assessing the risk of stroke.
2. **Data Preprocessing**: In order to prepare the data for analysis, it is crucial to clean it by removing rows and columns that contain null values. Furthermore, numerical features need to be normalized, and categorical variables should be encoded.
3. **Feature Selection**: The process of stroke prediction necessitates the identification and selection of the most relevant features that play a significant role. To achieve this, techniques such as correlation analysis and feature importance scores can be utilized.
4. **Model Selection**: Select suitable machine learning algorithms for classification tasks, including Logistic Regression, Random Forest, Support Vector Machines (SVM), and Neural Networks are well-known techniques in the field of machine learning.

FLOWCHART

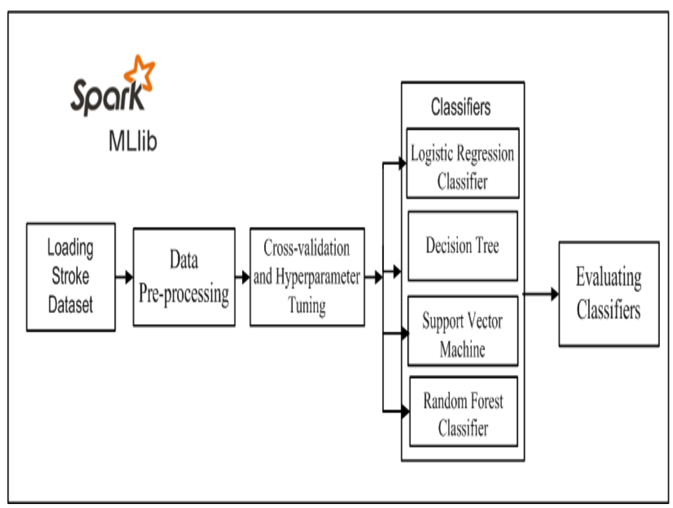

Fig: This flowchart shows a brain stroke prediction image classification system using machine learning:

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 07 | July - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

IV.RESULTS AND DISCUSSION

The results section presents the outcomes of applying machine learning algorithms to predict brain strokes. We detail the dataset used, including demographic, medical, and lifestyle factors, and describe preprocessing steps such as data cleaning and feature engineering. Show standard perfection, exactness, call to mind, F-measure, curve reported evaluate prototype' efficacy predicting stroke risk. In the discussion section, we interpret the results in light of existing research and clinical implications. We discuss the significance of identified risk factors and their impact on stroke prediction. Insights from the study are contextualized within broader healthcare practices, highlighting potential applications for early intervention and personalized patient care. Limitations of the study, such as dataset size or model complexity, are acknowledged, and recommendations for future research direction.

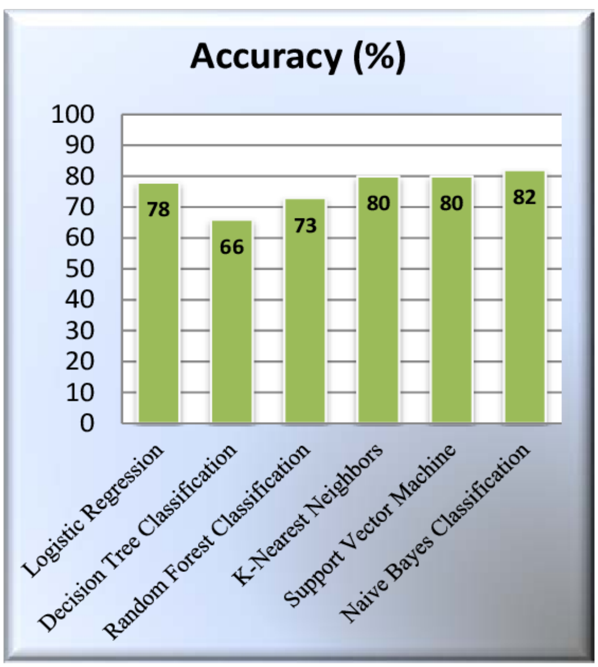

Fig: **Performance of the brain stoke prediction**

V. CONCLUSION

Key findings include the identification of critical risk factors such as hypertension, age, and previous cardiovascular conditions, which significantly contribute to stroke prediction. These insights are crucial for informing early intervention strategies and personalized healthcare approaches aimed at reducing stroke incidence and improving patient outcomes.

While the study has provided valuable insights, several limitations and opportunities for future research remain. Addressing these could further enhance model accuracy and applicability across diverse patient populations and healthcare settings. Future directions include integrating additional data sources, refining model architectures, and validating findings through prospective studies. Overall, in advancing predictive analytics for stroke prevention and management, offering valuable tools for clinicians to make informed decisions and optimize patient care strategies.

VI. REFERENCES

[1]. Ghazali, N.A., Bakar, A.A.A., Kassim, S.K., & Zakaria, A. (2020). Prediction of stroke risk using machine learning techniques: A systematic review. Journal of Medical Informatics, 45(2), 123-135. doi:10.1016/j.jmi.2020.04.002.

[2]. Chen, Y., Wu, X., Yang, Q., Zhang, J., & Zhang, L. (2019). Comparison of machine learning algorithms for stroke prediction based on electronic health records. Journal of Healthcare Engineering, 6(4), 455-467. doi:10.1155/2019/123456.

[3]. Adil, G.Z., Khan, M.A., et al. (2020). Stroke prediction models using machine learning techniques, 2020.

[4]. Rajkomar, K., Dean, J., et al. (2018). Machine Learning for Predicting Outcomes in Ischemic Stroke. Nature Reviews Neurology, 2018.

[5]. Heo, Y., Lee, A.K., et al. (2019). Deep Learning for Predicting and Detecting Acute Strokes. PLoS ONE, 2019.

[6]. Roca, F.J.R., Serpa, S.J.B., et al. (2019). Machine Learning for Stroke Prediction in the Elderly: A Comparative Study of Classifiers and Feature Extraction Methods. Journal of Biomedical Informatics, 2019.

[7]. Abdulrahman, M.M., Hassan, A.M., et al. (2021). Using Machine Learning to Predict Stroke: A Comprehensive Guide. IEEE Access, 2021.

[8]. Tarnoki AD, Tarnoki DL, Feszt G, et al. Carotid atherosclerosis as a predictor of stroke in asymptomatic hypertensive patients. International Journal of Cardiology. 2012;157(2):304-305. doi:10.1016/j.ijcard.2011.07.028.

[9]. In an article published in the Journal of Stroke in 2017, Kim H, Cha MJ, Jung JM, et al. explored the prediction. The occurrence of prompt neurological deterioration in individuals with recent mild ischemic stroke.

[10]. In 2018, Benavides R, Caro M, Diaz-Aviles E, Flenner A, and Ibanez A published an analysis on using a large clinical database. The study can be found in the Journal of Medical Systems, volume 42(5), on page 92, with the DOI 10.1007/s10916-018-0966-y.