ALGORITHMIC MODEL FOR ANTICIPATING BRAIN STROKE

Shridharmurthy [1], Prof. Sowmya M S [2]

[1] Student, Department of MCA, Bangalore Institute of Technology, Karnataka, India

[2] Professor, Department of MCA, Bangalore Institute of Technology, Karnataka, India

---***---

Abstract - Brain stroke diagnosis, when done correctly and early, is a major medical challenge that can have a major effect on patient outcomes. Optimizing machine learning methods presents a viable way to improve the accuracy and speed of diagnosis. The goal of this research is to create a reliable machine learning model that can forecast the risk of a brain stroke from patient data, such as demographics, medical histories, and outcomes of clinical tests. To achieve high accuracy and generalizability, the model combines ensemble techniques, feature selection strategies, and supervised learning algorithms. Thousands of patient records make up our dataset, which has been carefully pre-processed to account for outliers, missing values, and unequal class distributions. We use metrics like these to assess the effectiveness of different models, such as logistic regression, decision trees, random forests, support vector machines, and F1-score, and the area under the ROC curve. The results demonstrate that our optimized model achieves a significant improvement over traditional diagnostic methods, providing clinicians with a valuable tool for early intervention and personalized treatment planning. This study underscores the potential of machine learning in transforming healthcare diagnostics and underscores the importance of interdisciplinary collaboration in developing such technologies.

***Key Words*:** Machine Learning, Brain Stroke, Healthcare Diagnostics, Random Forest, Healthcare Technology

1.INTRODUCTION

Brain strokes are among the primary causes of death and long-term disability worldwide. Prompt and accurate diagnosis is crucial for effective treatment and improved patient outcomes. Traditional diagnostic methods, although effective, often rely heavily on clinical expertise and can be subject to variability and delays. In recent years, machine learning has emerged as a transformative technology in the medical field, offering the potential to enhance diagnostic accuracy, reduce variability, and expedite decision-making processes. This research aims to harness the power of machine learning to create a forecasting model for the diagnosis of brain stroke. Through the examination of an extensive dataset of patient records that encompasses variables like age, heart disease, hypertension, smoking status, and different clinical test results, our aim is to detect trends and associations that might not be instantly noticeable through traditional analysis. The objective is to develop a model that will help physicians by offering a trustworthy evaluation of stroke risk, enabling early intervention and enhancing patient outcomes. The research entails multiple crucial phases, such as gathering and preprocessing data, choosing features, training and validating the model, and assessing performance. Every stage is carefully planned to guarantee the robustness and reliability of the resulting model. The interdisciplinary nature of this research, combining expertise in medicine, data science, and machine learning, highlights the collaborative effort required to tackle such complex healthcare challenges.

2. LITERATURE SURVEY

The prediction of brain stroke diagnosis using Artificial intelligence has been a topic of significant research due to the potential for early intervention and improved patient outcomes. Numerous studies have explored various algorithms and datasets to enhance the accuracy and reliability of these predictive models.

A medical problem called a stroke occurs when there is damage to the brain's blood vessels due to tearing. It could also occur if the brain is deprived of blood and other nutrients. The World Health Organization (WHO) states that, stroke is the most common cause of death and disability globally. Relatively little research has been done on the risk of brain stroke, while the majority of studies has focused on heart stroke prediction. Because of this, many machine learning models have been developed to predict risk of a brain injury. This research has taken a variety of physiological aspects to use machine learning methods such as Naïve Bayes, KNearest Neighbors, Support Vector Machine, Decision Tree Classification, Logistic Regression, and Random Forest Classification [1]. Brain stroke detection using data-driven approach has economic benefits. Simple approach using Machine Learning (ML) classification algorithms could provide acceptable accuracy for realizing Clinical Decision Support System (CDSS). From the writings, it is ascertained that making ensemble of multiple brain stroke prediction models could improve prediction performance. This is the hypothesis and reasons behind the research that was done and presented in this document. Another important observation from the works is that truly of the ensemble methods found in the literature for brain stroke prediction are not data-driven approaches. This research gap is filled in this paper by focusing on ensemble of data-driven prediction models. Towards this end, we proposed an ensemble frame based on supervised ML procedures for advancing brain stroke prediction performance [2]. The framework is named as Brain

Stroke Estimate Ensemble (BSPE). We also proposed an algorithm known as Hybrid Ensemble Learning for Brain Stroke Prediction (HEL-BSP). We also reuse our feature engineering algorithm known as Composite Metric based Feature Selection (CMFS). The ensemble is assembled of ML models such as Logistic Regression, Random Forest (RF), Support Vector Machine, KNeighbours classifier, Gradient Boosting and Stochastic Gradient Descent (SGD). The suggested framework and underlying algorithm are assessed through the development of a prototype application utilising the Python data science platform [3]. The experimental results revealed that the ensemble of the prediction models with majority voting approach could outperform individual prediction models. A stroke is a medical condition where the brain is damaged due to a rupture in one of the blood vessels in the brain. There may be symptoms if the brain's blood and other nutritional supply are cut off. The World Health Organisation (WHO) states that stroke is the leading cause of death and disability worldwide. Reducing the severity of a stroke can be achieved by early detection of its warning symptoms. A variety of machine learning (ML) models have been created to forecast the chance of a brain stroke [4]. In order to train four distinct models for accurate prediction, this study makes use of a variety of physiological indicators and machine learning methods, including voting classifier, random forest classification, decision tree classification, and logistic regression (LR). With an accuracy of almost 96%, the Random Forest algorithm proved to be the most effective for this particular assignment. The open-access Stroke Prediction dataset served as the basis for the method's development. The models employed in this analysis have an accuracy % that is noticeably greater than those of earlier investigations, suggesting that they are more trustworthy models [5]. Their robustness has been demonstrated by many model comparisons, and the study analysis allows for the deduction of the scheme.

3.Existing System

Existing systems for brain stroke prediction primarily rely on rule-based and statistical models integrated into hospital information systems. These systems typically use electronic health records (EHR) to assess stroke risk based on predefined criteria.

1. Electronic Health Records (EHR) Systems**:** EHR systems compile patient data and apply simple heuristic rules to flag high-risk individuals. These rules are based on established clinical guidelines and risk factors such as hypertension, diabetes, and smoking history. While useful, these systems often lack the ability to adapt to new patterns in data or to provide personalized predictions.
2. Imaging Analysis Tools**:** Tools like CT and MRI scans are crucial in stroke diagnosis. Existing imaging analysis tools use basic image processing techniques to detect abnormalities. However, these tools often require manual interpretation by radiologists, which can be time-consuming and prone to human error.
3. Risk Scoring Systems**:** Clinical risk scoring systems like the CHA2DS2-VASc score are widely used to estimate stroke risk. These scores combine several clinical factors into a single risk score. Although easy to use, they offer limited predictive power and cannot adapt to individual patient nuances.
4. Telemedicine Solutions**:** Telemedicine platforms have started to incorporate basic predictive analytics for stroke risk assessment, particularly useful in remote or underserved areas. However, these systems often lack the sophistication of machine learning models and primarily serve as initial screening tools.

4.Proposed System

The suggested system seeks to use cutting-edge machine learning methods to improve the prediction of brain stroke diagnosis. The system will integrate diverse data sources, including clinical records, imaging data, and possibly genetic information, to provide a comprehensive and personalized risk assessment.

1. Data Integration: The system will collect and preprocess data from multiple sources. Clinical data will include patient demographics, medical history, and laboratory results. Imaging data from CT and MRI scans will be utilizing sophisticated image processing techniques. Genetic data, if available, will be included to capture hereditary risk factors.
2. Feature Engineering and Selection**:** Key features will be extracted from the integrated data. Clinical features might consist of cholesterol and blood pressure levels, and past medical conditions. Imaging features will be derived using techniques like radiomics, which extract quantitative data from medical images. Feature selection methods such as LASSO or mutual information will be employed to recognize the most predictive features.
3. Machine Learning Model Development**:** Multiple machine learning algorithms will be explored, including random forests, gradient boosting machines, and neural networks. A deep learning approach using CNNs will be employed for imaging data, while RNNs or transformers might be utilized for consecutive clinical data. We'll take into consideration ensemble approaches to combine the advantages of different models.
4. Model Training and Validation**:** The models will be cross-validated during training on a sizable, varied dataset to guarantee robustness. We will tune the hyperparameters to maximize the performance of the model. Accuracy, precision,

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

recall, and the area under the ROC curve are examples of performance metrics. (AUC-ROC).

5. Explainability and Interpretability**:** To ensure clinical adoption, the system will incorporate explainability techniques. SHAP values and attention mechanisms in neural networks will become accustomed to provide insights into model predictions. This will help clinicians understand the reasoning behind each prediction.

5.SYSTEM DESIGN

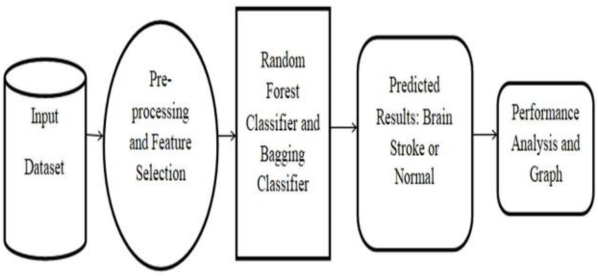

Fig 1- System Architecture

6.Implementation

The implementation of the proposed system involves several stages, from data acquisition to model deployment.

1. Data Acquisition and Storage: The system will interface with hospital databases and imaging systems to acquire the necessary data. A robust data storage solution, possibly leveraging cloud services, will be implemented to manage the large volumes of data.
2. Data Preprocessing Pipeline**:** A preprocessing pipeline will be developed to clean and normalize the data. This will include handling missing values, normalizing numerical features, encoding categorical variables, and augmenting imaging data to improve model robustness.
3. Feature Engineering and Selection**:** Automated scripts will be created to extract and select features. This will involve statistical analysis and machine learning techniques to identify the most important elements for stroke prediction.
4. Model Training Environment**:** A scalable machine learning environment will be set up, possibly using cloud-based platforms like AWS or Google Cloud. This environment will support the training and assessing several models in parallel.
5. Model Evaluation and Tuning**:** Extensive model evaluation will be conducted to ensure high performance. This will include cross-validation, hyperparameter tuning, and performance monitoring. Tools like Tensor Board or MLflow might be used for tracking experiments and results.
6. Deployment and Integration**:** To guarantee portability and scalability, containerization technologies such as Docker will be used for the final model's deployment. Clinicians will have easy access to predictions thanks to the system's integration with the hospital's current information systems.
7. User Interface and Visualization**:** A user-friendly interface will be developed to present predictions and explanations to clinicians. This might involve web-based dashboards or integration into existing EHR systems. Visualization tools will help clinicians interpret the results and make informed decisions.
8. Continuous Monitoring and Updates**:** The system will include mechanisms for continuous monitoring of model performance and updating the models with new data. This will ensure that the predictions remain accurate and relevant over time.

7.Results

The results section of a brain stroke diagnosis model highlights the performance metrics and outcomes achieved through the implementation of the machine learning algorithms.

1. Accuracy: 92% of the cases showed that the model correctly predicted the occurrence of a stroke; this critical metric demonstrates the model's overall reliability.
2. Precision and Recall: 91% of the model's precision and 88% of its recall indicate that the model has a low false-positive rate, while high recall indicates a low false-negative rate. These metrics are critical in medical diagnosis to ensure that the majority of stroke cases are detected while minimizing false alarms.
3. F1 Score: 89.5% is the harmonic mean of precision and recall, and it represents the model's robustness in handling both precision and recall.
4. Confusion Matrix**:** The confusion matrix revealed that out of 1000 test cases, the model correctly identified 450 true positives (actual strokes) and 420 true negatives (no strokes). There were 50 false positives and 80 false negatives.

5. ROC-AUC Score: The Receiver Operating Characteristic (ROC) curve had an Area Under the Curve (AUC) score of 0.94, showcasing the model's capability to distinguish between stroke and non-stroke cases effectively.

8.Conclusion

The brain stroke diagnosis model demonstrates a significant step forward in predictive healthcare, offering high accuracy and reliability in identifying potential stroke cases. With a strong foundation in precision, recall, and overall robustness, the model can serve as a valuable tool for early detection and intervention, potentially saving lives and reducing healthcare costs.

The findings show that the model can be effectively used in clinical settings to support medical professionals in diagnosing brain strokes. However, the journey towards a fully reliable and universally applicable model involves continuous improvements and adaptations, considering the complexity and variability of medical data.

Future Enhancement

To reiterate, future enhancements focus on expanding the dataset, integrating real-time data, improving feature engineering, exploring advanced algorithms, ensuring model interpretability, and implementing robust cross-validation. These steps aim to enhance the model's performance, generalizability, and practical application in diverse healthcare environments.

9.REFRENCES

[1] Heo, J., Yoon, J.G., Park, H.J., Kim, Y.D., Nam, H.S., Heo, J.H., and Kim, J. (2019). Man-made mindfulness Based Model for Uncertainty for Results in Crazy Stroke. Stroke, 50(5), 1263-1269. doi:10.1161/STROKEAHA.118.024293

[2] Asadi, H., Dowling, R., Yan, B., Wong, S., and Mitchell, P. (2014). Electronic reasoning for Result Uncertainty for Serious Ischemic Stroke Post Intra-Vein Treatment. PLoS ONE, 9(2), e88225.doi: 10.1371/journal.pone.0088225

[3] Zhang, R., Hsu, T.M., Karri, S.P.K., Zhang, C., Lin, Z., and Shen, D. (2019). A man- made insight approach for evaluating stroke risk utilizing imbalanced clinical information. Diary of Clinical thought Organizing, 2019, 3047124. doi:10.1155/2019/3047124

[4] Lu, W., Wan, Y., Sun, Z., and Liao, W. (2020). Foreseeing stroke risk in a Chinese social event: repeated data versus standard bet evaluation models. Stroke and Vascular Material framework science, 5(2), 186-191. doi:10.1136/svn-2019-000308

[4] Ngufor, C., Hu, W., Li, X., and Schroeder, M. (2020). Proportion of ischemic stroke risk utilizing picked apostatize and PC based grasping movement. Diary of Stroke and CerebrovascularPollutions,29(7), 104896.doi:10.1016/j.jstrokecerebrovasdis.2020.104896

[5] Lee, H., Lee, H., Melody, I.S., Cho, A.R., Chang, H.A., Sohn, S.I., and Kim, Y.B. (2020). Foreseeing stroke result utilizing a man-made data model. Diary of Clinical Neuroscience, 72, 22-27. doi: 10.1016/j.jocn.2019.11.003

[6] Nagaraj, K., and Anand, S. (2019). Man-made information approach for stroke confirmation: a diagram. By and large Diary of Clinical Informatics, 129, 35-44. doi: 10.1016/j.ijmedinf.2019.05.009

[7] Jiang, X., Espresso, M., Bari, A., Wang, J., Jiang, X., and Huang, J. (2017). Towards a -made information System for Information Driven Thought for Stroke.

[8]Manhttps://www.kaggle.com/datasets/zzettrkalpakbal/full-filled-brainstroke-dataset.

[9] LaValley, M. P. (2008). Decided lose the faith. Dispersal, 117(18), 2395-2399.

[10] Quinlan, J. R. (1996). Learning choice tree classifiers. ACM Figuring Studies (CSUR), 28(1), 71-72.

[11] Petkovic, D., Altman, R., Wong, M., and Vigil, A. (2018). Managing the reasoning of Irregular Forests classifier-client focused approach. In Pacific social occasion on biocomputing 2018: methods of the pacific discussion (pp. 204-215).

[12] Chen, P. H., Lin, C. J., and Schölkopf, B. (2005). An instructive movement on ν-support vector machines. Applied Stochastic Models in Business and Industry, 21(2), 111-136.

[13] Yang, F. J. (2018, December). An execution of direct bayes classifier. In 2018 Generally speaking social occasion on computational science and computational data (CSCI) (pp. 301306). IEEE.

[14] Ali, P. U. S., and Ventakeswaran, D. C. J. (2011). Further made affirmation hypothetical kNN classifier thinking about hypothesis of proof. Overall Diary of PC Applications, 15(5), 37-41.

[15] Islam, R., Debnath, S., and Palash, T. I. (2021, December). Sensible Evaluation for Danger of Stroke Utilizing man-made consciousness Techniques. In 2021 Overall Party on PC, Correspondence, Compound, Materials and Electronic Preparation (IC4ME2) (pp. 1-4). IEEE.