International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Brain Stroke Prediction Utilizing Convolutional Neural Networks

Mrs Jhansi Rani
Asst.Professor, Department of Computer Science and Engineering(AI&ML), Sreenidhi Institute of Science and Technology

Virija Vidapu
Department of Computer Science and Engineering(AI&ML), Sreenidhi Institute of Science and Technology

Sunkara Sahrudai
Department of Computer Science and Engineering(AI&ML), Sreenidhi Institute of Science and Technology

Syed Omar Farooq
Department of Computer Science and Engineering(AI&ML), Sreenidhi Institute of Science and Technology

Abstract - A brain stroke is a serious medical illness that can have life-threatening consequences and is caused by to interruption in blood flow to the brain. Early detection and timely treatment can reduce mortality rates and also enhance patient recovery. Machine learning techniques have evolved into highly effective tools in recent years for predicting the risk of a brain stroke based on many health indicators. With a focus on important risk factors like age, blood glucose levels, heart disease, smoking, and hypertension, this study investigates the creation of a predictive model for brain stroke using convolutional neural networks. The goal of the model is to accurately classify individuals who are at risk by using a dataset that contains key medical information. The suggested method can help medical practitioners by facilitating early detection and preventative actions that minimize the impact of brain strokes.

Keywords - *Brain stroke prediction, Machine Learning, Predictive model, Convolutional Neural Networks (CNNs), Healthcare Risk Assessment.*

I. INTRODUCTION

Brain stroke is a severe medical condition that occurs when the blood supply to a part of the brain is disrupted, causing impairment, disability, or even death if not treated within time. Strokes are classified into two main types: **ischemic stroke** and **hemorrhagic stroke**.

Ischemic Stroke – This is the most common type of stroke, which is seen in almost 87 % of stroke cases. It occurs when a blood clot builds up and blocks the blood flow in the brain's arteries, leading to oxygen deprivation in affected areas.

Hemorrhagic Stroke – This occurs when a weakened blood vessel ruptures, causing bleeding in or around the brain. It is often linked to conditions such as high blood pressure. This type of stroke is less common than ischemic stroke cases and is seen in patients with high blood pressure.

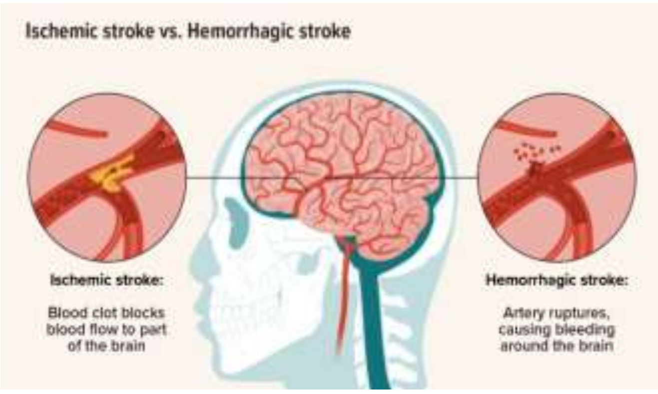

Fig 1.1 Ischemic vs hemorrhagic stroke

In this work, we use convolutional neural networks (CNN) to construct a predictive model for the detection of brain stroke. Our approach aims to identify precisely the individuals that could be at risk of having a brain stroke by taking key health indicators like age, blood sugar levels, hypertension, history of heart disease, smoking habits, etc, into consideration. The interpretability and robustness of the CNN-based model provide healthcare professionals with valuable insights into the relative importance of each risk factor.

This, in turn, facilitates more informed clinical decisions and targeted interventions, potentially reducing the incidence of severe stroke. The study findings show the potential of integrating artificial intelligence into routine medical practice, paving the way for more proactive and personalized healthcare solutions.

II. RELATED WORK

In recent years, various research efforts have explored the application of machine learning (ML) and deep learning techniques for predicting brain stroke, primarily using structured, non-imaging datasets composed of clinical and demographic information.

Wahed, Fahad, and Ahmed (2024), in their research titled " Prediction of Brain stroke using Machine learning algorithms and deep neural networks, concentrated on specific structured datasets obtained from platforms such as Kaggle. The authors assessed various algorithms, including randomforest, XGBoost, AdaBoost, Logistic regression, and deep neural networks. Notably, the random forest model achieved the highest prediction accuracy of 99%, while a four-layer neural network reached 92.39%. They concluded that the traditional machine learning models surpassed deep neural networks in efficiency and accuracy for stroke prediction when dealing with structured data [1].

Mehta, Adhikari, and Sharma (2021), in their thorough review "Machine Learning Techniques in Stroke Prediction," published in the Journal of Medical Systems, explored a broad spectrum of algorithms, including decision trees, random forests, and neural networks, for predicting stroke risk. Their examination concentrated on how these models interpret structured data, such as age, hypertension, smoking habits, and glucose levels, to estimate stroke risk factors and outcomes. This research highlights the continuous advancements in employing traditional machine learning approaches to improve the early detection of strokes [2].

In contrast to earlier research by Thompson et al.(2019) and Wahed et al. (2020), which focused on stroke classification using CNNs and medical imaging data, the present study aims to predict stroke using clinical and demographic data of the patient rather than relying on image-based diagnostic methods. This differentiation is significant, as models based on structured data are more readily accessible and scalable within real-world health care settings, particularly for earlier risk assessment.

Together, these studies emphasize the increasing potential of machine learning and deep learning methods in stroke prediction, particularly when applied to non-imaging clinical data. They also lay the groundwork for the current research, which intends to use Convolutional Neural Networks (CNNs) for predictive modeling utilizing structured health information rather than medical images.

III. METHODOLOGY

A.Data collection

The dataset used in this study was obtained from a publicly available source on Kaggle, containing **4,981 records** with **10 attributes** relevant to brain stroke prediction.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

These attributes include key risk factors such as age, hypertension, heart disease, blood glucose levels, smoking status, and body mass index (BMI), among others.

The dataset provides a diverse range of patient information, enabling the development of an effective predictive model. The data was preprocessed to handle any inconsistencies and ensure its suitability for training and evaluating the proposed Convolutional Neural Network (CNN)-based model.

B.Data preprocessing

Data preprocessing plays a crucial role in predictive models. To ensure accurate results, the model must be trained on good-quality data that does not have outliers or missing values. The dataset used in this study was already preprocessed, with no missing values or outliers present. As a result, no additional data cleaning or transformation steps were required. The data was directly used for training and evaluation, ensuring a streamlined predictive model implementation.

1. Addressing the class imbalance

An imbalance between two classes exists when one class has way more samples than the other class, which will lead to biased predictions. In this case, we have 4,733 non-stroke cases and 248 stroke cases. We can see the imbalance between the 2 cases, which will lead to inaccurate predictions, so it was resolved using SMOTE (Synthetic Minority Oversampling Technique). SMOTE is an oversampling technique that generates synthetic samples for the minority class by selecting a random instance from the underrepresented category. It then identifies the **k-nearest neighbors** of the chosen instance within the feature space. A new synthetic data point is created by interpolating between the selected sample and one of its nearest neighbors. This interpolation follows the formula:

$$X_{new} = X + \lambda * (X_{nn} - X)$$

where X is the random sample selected from the minority class, λ is a randomly chosen value between 0 and 1, and X_{nn} is one of the selected k-nearest neighbours.

SMOTE ensures that the new sample lies somewhere along the line segment connecting the original data point and its neighbor, effectively increasing the diversity of the minority class without merely duplicating existing instances.

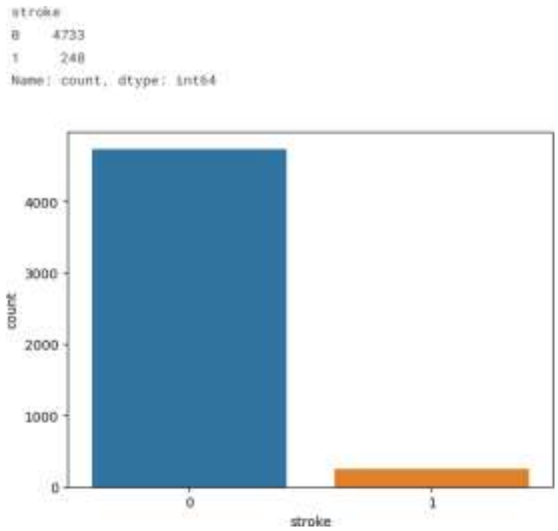

Fig 3.1.1 Data imbalance in classes before applying SMOTE

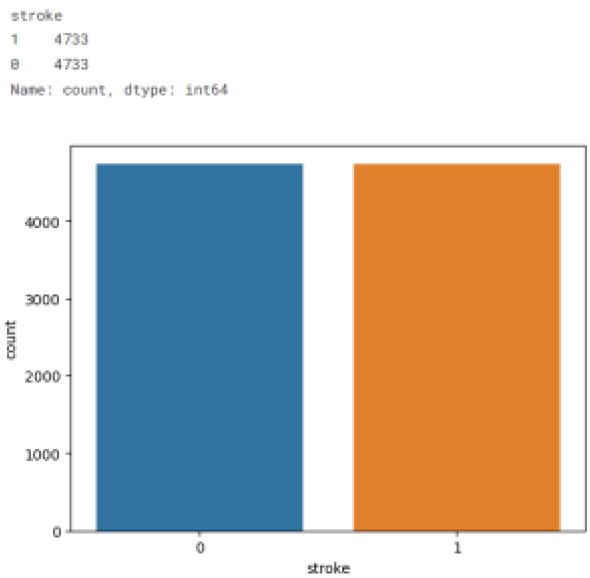

Fig 3.1.2 Balanced classes in the dataset using SMOTE

2. Data Splitting and Model Development

In this study, we employ a Convolutional Neural Network (CNN) for predictive analysis. The dataset is first divided into training and testing subsets using an 80:20 split ratio, ensuring that the model is trained on a diverse set of data while maintaining a portion for evaluation. The input data consists of 10 features representing relevant medical parameters. To facilitate training, the dataset is reshaped to be compatible with the CNN

architecture by adding a dimension, transforming it into a format suitable for 1D convolutional layers.

The CNN model is constructed using the **Sequential API** in TensorFlow, incorporating the following layers:

First Convolutional Layer: A 1D convolutional layer with 32 filters, a kernel size of 3, and a ReLU activation function. This layer extracts essential features from the input data. A max-pooling operation is applied to reduce dimensionality and improve computational efficiency.

Second Convolutional Layer: A similar 1D convolutional layer with 64 filters, also followed by a max-pooling operation to further refine feature extraction.

Flattening Layer: The extracted features are flattened into a one-dimensional vector before passing to the fully connected layers.

Fully Connected Layers: A dense layer with 128 neurons and an ReLU activation function is used to process extracted features. A dropout layer with a rate of 0.5 is included to prevent overfitting.

Output Layer: A single neuron with a sigmoid activation function is utilized for binary classification, predicting the likelihood of an individual being at risk.

The model is compiled using the Adamax optimizer with a learning rate of 0.001 and a binary cross-entropy loss function, given the binary nature of the classification task. The model is trained for 1000 epochs with a batch size of 128, using validation data to monitor performance.

This structured approach ensures an effective deep learning model that can identify patterns in medical data, aiding in predictive healthcare solutions.

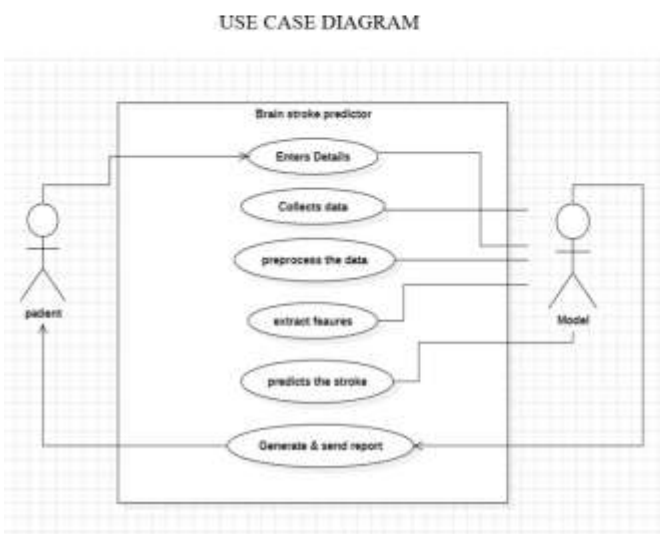

Fig 3.2.1 Use case diagram for brain stroke predictor

3. Performance metrics

In this study, to evaluate the effectiveness of the predictive model, we used **accuracy and test loss** as the two key performance metrics that evaluate the model's performance.

Accuracy is a commonly used metric in classification problems that measures how well a model correctly predicts outcomes. It is defined as the ratio of correctly classified instances to the total number of predictions. The formula for accuracy is:

Accuracy = Correct predictions / Total predictions

Accuracy is simply achieved by predicting the majority class while failing to identify the minority class correctly.

Hence, accuracy is often evaluated alongside other metrics to get a more comprehensive understanding of the model's performance. Another way to represent accuracy, particularly in classification problems, is through the following formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where:

TP (True Positives) – Instances correctly predicted as positive.

TN (True Negatives) – Instances correctly predicted as unfavorable.

FP (False Positives) – Instances incorrectly predicted as positive.

FN (False Negatives) – Instances incorrectly predicted as unfavorable.

Test Loss

Test loss represents how well the model generalizes the unseen data. It is computed using a loss function, which quantifies the difference between the predicted and actual values.

A lower test loss value indicates that the model's predictions closely align with actual outcomes, whereas a higher loss suggests poor generalization. Unlike accuracy, test loss provides a more detailed understanding of model performance, helping identify potential overfitting or underfitting issues. By analyzing both accuracy and test loss, we can assess the model's predictive strength and ensure its reliability in real-world applications.

IV REFERENCES

1."Prediction of Brain Stroke using Machine Learning Algorithms and Deep Neural Network Techniques" by Wahed, Fahad, and Ahmed (2024)

2."Machine Learning Techniques in Stroke Prediction" by Mehta, Adhikari, and Sharma (2021)

3. "Effective Brain Stroke Prediction with Deep Learning Model by Incorporating YOLO_5 and SSD." by Sailaja, Y., & Pattani, V. (2023)

4. "Fusing Attention and Convolution: A Hybrid Model for Brain Stroke Prediction." by Bhuvanya, R., Kujani, T., & Sivakumar, K. (2024)

5. "Acute ischemic stroke lesion segmentation in non-contrast CT images using 3D convolutional neural networks." by Dobshik, A. V., et al. (2023)

6. "CNN-LSTM Based Multimodal MRI and Clinical Data Fusion for Predicting Functional Outcome in Stroke Patients." by Hatami (2022)

7. "Early-stage stroke prediction based on Parkinson and wrinkles using deep learning." by Haritha, T., & Babu, A. V. S. (2024)

8. "Brain Stroke Prediction Using Deep Learning: A CNN Approach." by Reddy, M. K.

9. "Deep Learning–Based Brain Computed Tomography Image Classification with Hyperparameter Optimization through Transfer Learning for Stroke." by Chen, Y(2022)

10. "Artificial intelligence-based identification of brain CT medical images." by Chen, R. (2022)

11. "Predictive Analysis for Risk of Stroke Using Machine Learning ." by Islam, R. (2021)

12. "Deep Learning Enhanced Internet of Medical Things to Analyze Brain Computed Tomography Images of Stroke Patients," by Omarov, B. (2023)

13. "Deep Learning-Enhanced Internet of Medical Things to Analyze Brain CT Scans of Hemorrhagic Stroke Patients: A New Approach." by Xu, (2021)

14. "Performance evaluation of machine learning classifiers for brain stroke prediction." by Arora, D. (2024)

15. "Brain Stroke Prediction Using Ensemble Learning." by Krishna, N. G., et al. (2024)

16. "Integrating uncertainty in deep neural networks for MRI-based stroke analysis." by Herzog, L., et al. (2020)

17. "Automatic Post-Stroke Lesion Segmentation on MR Images using 3D Residual Convolutional Neural Network." by Tomita, N., et al. (2019)

18. "Stroke Risk Prediction Using Machine Learning Algorithms." by Gurjar, R., et al. (2022).

19. "Analyzing the performance of stroke prediction using ML classification algorithms." by Sailasya, G., & Kumari, G. L. A. (2021)

20. "Stroke prediction through Data Science and Machine Learning Algorithms." Tavares, J.-A. (2021)

V RESULTS

The performance of the proposed CNN model for brain stroke prediction was evaluated using accuracy and test loss. The model achieved a **highest accuracy of 97.2%** during training and a test accuracy of 96.11%, demonstrating its strong predictive capability. Additionally, the test loss was recorded at 0.1260, indicating minimal error in predictions. These results highlight the model's ability to generalize well to unseen data while maintaining high accuracy. The combination of low test loss and high accuracy confirms the model's reliability in identifying individuals at risk of stroke, making it a promising tool for medical diagnostics.