CerebroIntellex Leveraging Deep Learning Framework for Stroke Analysis

S. Thenmalar[1], Vansh Sharma[2], Divija Agrawal[2] and Shubhangi Pandey[2]
[1]*Department of Networking and Communications, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Chennai – 603203, Tamil Nadu, India*
[2]*Department of Networking and Communication, SRM Institute of Science and Technology, Kattankulathur, Chennai – 603203, Tamil Nadu, India*

Keywords: Stroke Prediction, Linear Discriminant Analysis, Convolutional Neural Networks, Ischemic Stroke, Hemorrhagic Stroke, Machine Learning, Medical Imaging.

Abstract: Stroke is still the leading cause of morbidity and mortality in many parts of the world, therefore, early prediction and classification need to be developed. This work proposes a two-tiered stroke detection and classification approach using machine learning and deep learning models. In the first stage, it is a model using a questionnaire based on LDA that predicts the probability of stroke occurrence through analysis of important risk factors, such as age, BMI, average glucose level, hypertension, and lifestyle attributes. The second stage uses datasets from CT and MRI images for classification into ischemic, hemorrhagic, or normal cases using CNN. The proposed system aids in the early detection of strokes with enhancing diagnostic accuracy in using multimodal data. Comprehensive evaluations demonstrate high prediction accuracy and robust classification performance. This contributes to personalized healthcare by developing risk factor analysis combined with imaging techniques and provides a scalable solution for clinical application.

1 INTRODUCTION

Stroke is among the most prominent causes of morbidity and mortality globally, having a severe influence on global health systems and individual lives. Wolfe et al. highlighted the ruinous consequences of stroke on individuals as well as society, pointing towards the necessity for better prevention and intervention measures. The Global Burden of Disease Study (GBD 2019 Stroke Collaborators) also highlighted the rising trend of stroke, determining the most significant risk factors that have fueled its burden over the last three decades. Balakrishnan et al. responded to these issues by emphasizing the need for early detection of stroke and accurate classification in enhancing patient outcomes through timely and effective medical interventions.

Recent developments in machine learning and deep learning have greatly enhanced stroke prediction and diagnosis. Elsaid et al. and Alanazi et al. proved the capability of ML models to improve the early detection of stroke, especially by processing patient-specific clinical and imaging data. Zhang et al. and Fernandes et al. investigated the use of ML algorithms for predicting stroke risk, with a illustration of how parameters like age, blood pressure, and lifestyle play a role in determining stroke likelihood. Supervised machine learning algorithms have been instrumental in prediction models, as Hassan et al. revealed their usefulness in the analysis of stroke risk and determining factors that contribute to it. Aside from that, deep learning methods are also shown to be extremely capable of distinguishing between ischemic and hemorrhagic strokes based on medical imaging information (Subudhi et al.; Sailasya et al.) for more accurate and automatic diagnosis. The paper combines machine learning and deep learning methods in an integrated approach towards stroke prediction and classification.

Two main components are the system suggested by this proposed system. The initial portion utilizes predictive analysis based on supervised learning models such as Linear Discriminant Analysis (LDA) to analyze the risk of stroke according to responses of patients from a well-defined questionnaire (Sundaram et al.; Singh et al.). The second half includes a doctor's dashboard utilizing convolutional neural networks (CNNs) for identifying stroke

Thenmalar, S., Sharma, V., Agrawal, D. and Pandey, S.
CerebroIntellex Leveraging Deep Learning Framework for Stroke Analysis.
DOI: 10.5220/0013891300004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 3*, pages 44-54
ISBN: 978-989-758-777-1

categories from multimodal MRI and CT scan imaging inputs (Srinivas et al.; Sawan et al.).

Earlier research has highlighted the importance of hybrid models that merge clinical and imaging information to enhance stroke prediction and classification correctness. Rahman et al. and Lavanya et al. investigated the performance of such methods, showing that the incorporation of multiple data sources can greatly boost diagnosis. In an attempt to enhance the accuracy of detection, Gheibi et al. and Qasrawi et al. examined the application of ensemble classifiers and segmentation methods, which have found broad acceptance for automatic stroke detection. These improvements notwithstanding, the majority of models treat stroke prediction and classification independently, and therefore leave a gap in clinical decision-making. This research is responding to this challenge through proposing an integrated system that synthesizes questionnaire risk estimation with imaging classification (Abbasi et al.; Adam et al.).

With the help of recent developments in machine learning, deep learning, and multimodal data fusion, this study has set out to create a clinically significant and scalable solution for the management of stroke. The suggested system is to be used in aiding early stroke diagnosis, tailored treatment planning, and enhanced patient outcomes, all leading to supporting healthcare professionals to make quicker and more precise clinical decisions (Sharma et al.; Tegistu et al.).

2 RELATED WORKS

Stroke is one of the leading causes of death and disability worldwide, necessitating robust diagnostic tools and predictive frameworks. Wolfe et al. highlighted the burden of stroke on global health, emphasizing its devastating impact on individuals and societies. The Global Burden of Disease Study (GBD 2019 Stroke Collaborators) reported a significant increase in stroke cases over the past three decades, identifying key risk factors and advocating for enhanced diagnostic and predictive solutions.

Machine learning (ML) methods have been used widely for stroke prediction and classification. Balakrishnan et al. proved the efficiency of ML models in the early identification of hemorrhagic stroke using clinical and imaging information. Elsaid et al. proposed an ML model to predict hemorrhagic transformation based on the interaction of clinical predictors with imaging biomarkers. Likewise, Alanazi et al. used ML algorithms to forecast stroke risk from laboratory test results, emphasizing the role of electronic health records (EHRs) in the identification of high-risk patients.

Deep learning (DL) techniques have become increasingly popular in handling complex medical data. Zhang et al. proposed a deep learning model that can identify stroke lesions from MRI images with high segmentation accuracy. Hassan et al. highlighted the significance of feature engineering and predictive modeling in the identification of important stroke risk factors. Subudhi et al. presented an overview of different ML-based methods for ischemic stroke characterization based on MRI, with the focus on the application of deep learning in creating sophisticated imaging-based diagnostic tools.

Hybrid frameworks combining clinical and imaging information have proven to be very successful. Sailasya et al. compared various ML classification algorithms and determined ensemble algorithms to be best suited for predicting stroke. Singh et al. and Srinivas et al. used supervised learning-based strategies to classify strokes and showed pragmatic usage in clinical practice. Sawan et al. introduced a soft voting-based ensemble classifier using several algorithms to improve classification performance. Recent developments in DL architectures have further improved stroke detection. Lavanya et al. proposed a predictive model that integrated machine learning, clinical information, and sophisticated algorithms for early stroke detection. Gheibi et al. designed CNN-Res, a deep learning algorithm intended for segmenting acute ischemic stroke lesions from multimodal MRI images, showcasing how DL enhances lesion segmentation. Fernandes et al. presented a comprehensive review of machine and deep learning methods, their clinical usage, and problems in stroke diagnosis.

A few studies have addressed improving the precision of ischemic stroke detection with hybrid methods. Qasrawi et al. proposed a hybrid ensemble deep learning model with higher accuracy for ischemic stroke classification. Abbasi et al. surveyed deep learning models for ischemic stroke segmentation automatically, highlighting how they can be used in clinical environments. Adam et al. and Sharma et al. discussed supervised models for predicting stroke, highlighting the importance of multiple data sources in enhancing the accuracy of predictions. Predictive models focused on the patient have also been on the spotlight in recent years. Tegistu et al. proposed a deep neural network (DNN)-based model for stroke risk prediction from patient data, with notable improvements in early prognosis. Rahman et al. generalized this model to MRI-based stroke prediction, promoting the combination of imaging data with deep learning methods. Sirsat et al. and Hosseini et al. surveyed several ML and DL algorithms for stroke diagnosis, pointing out scalability and performance gaps.

There have been several review articles on the use of ML and DL in stroke prediction. Sensors (2024) gave an in-depth review of DL methods for diagnosing brain stroke and their potential for clinical use. IEEE Xplore (2023) and other research works (Frontiers in Neurology, Liu et al.) stressed the need to integrate electronic health records, medical images, and demographic data to make more accurate stroke predictions.

The expanding literature on hybrid models, ensemble classifiers, and deep learning architectures (Dev et al., Panachakel et al., Sundaram et al.) has provided the background for sophisticated diagnostic tools. By drawing from such earlier studies, this research proposes a consolidated dashboard-based method that combines ML and DL methods to forecast the threat of a stroke and diagnose ischemic and hemorrhagic strokes. This study fills the gap by integrating questionnaire-based risk assessment and MRI-based stroke categorization, eventually enhancing early detection, clinical decision-making, and scalability in stroke diagnosis and management.

3 METHODOLOGY

This research is conducted following a systematic methodology to predict the risk of stroke based on data from a questionnaire. The process is divided into five stages: data preprocessing, exploratory data analysis (EDA), feature selection, model development and evaluation, and deployment. Each stage is designed meticulously to ensure accuracy and interpretability in stroke prediction.

3.1 Questionnaire-Based Risk Assessment

3.1.1 Data Preprocessing

The sources used for data preparation were healthcare repositories which contained such data as age, BMI, average glucose levels, hypertension, heart disease, marital status, work type, residence type, and smoking status. The additional target variable is "stroke," signifying whether a patient has had a stroke. Columns like 'id' are superfluous, so these columns were dropped as well for redundancy. There were certain features that included missing values. The statistical imputation methods, especially mean replacement for missing entries, were applied. This would fill the missing entries with the mean of the respective column without creating any bias in the data. The dataset was checked for duplicate rows, which were removed to prevent data repetition. Extreme outliers in numerical features, such as BMI and glucose levels, were also handled using IQR methods to maintain data integrity. Categorical variables such as gender, smoking status, and work type were converted to numerical values. For instance, 'Yes' and 'No' for hypertension or heart disease were substituted with binary values (1 and 0) for compatibility with machine learning algorithms.

3.1.2 Exploratory Data Analysis (EDA)

The data distribution showed that the target variable was imbalanced, as stroke cases constituted less than 10% of the total data. Distribution of numerical features was age, BMI, and glucose. The result came out as: the greater these values are the more likelihood one has for getting a stroke.

- **Gender:** males 5.8%, female 4.7%
- **Hypertension:** if someone has hypertension chances of having stroke are 13.8%, for those without, it is only 3.7%.
- **Heart Disease:** A striking 16.4% of heart disease patients reported having had a stroke, as opposed to just 4.2% among those without.
- **Smoking:** Patients who smoked carried an 8.2% risk of having a stroke, whereas non-smokers carried only 4.1%.

Some of the features were found to be very influential in determining stroke risk, namely heart disease, hypertension, and smoking status. Histograms and bar charts were used to illustrate these facts.

3.1.3 Feature Selection

A correlation matrix was derived to establish how the numerical features, including age, BMI, and glucose level, could be related with stroke. Here, age was more strongly positively correlated with stroke while glucose level indicated a positive but lesser correlation. Applying mutual information scores on the categorical features assessed their predictiveness. The key contributors to the target variable have been identified, which include the smoking status as well as type of work.

Numerical: Age, BMI, Average Glucose Level.
Categorical: Hypertension, Heart Disease, Marital Status, Work Type, Residence Type, Smoking Status.

3.1.4 Model Development and Evaluation

Linear Discriminant Analysis (LDA) was used because it is suitable for multivariate data and performs well in binary classification problems such as stroke prediction. The dataset was divided into training (80%) and testing (20%) subsets. LDA model training was performed using the training set and validated on the test set.

- **Confusion Matrix:** The confusion matrix revealed that the classification was 85%, with true positives at 78%.
- **Precision-Recall Curve**: The balance of precision and recall showed high ability to detect stroke cases as the precision value was 82%.
- **ROC Curve and AUC:** The model obtained an AUC value of 0.89, which represents excellent performance in differentiating between cases and controls.

The learning curve demonstrated steady training and testing performances without much overfitting.

3.1.5 Deploy

A user-friendly application was developed to predict stroke risk based on patient inputs. The interface allows users to input key details such as age, BMI, glucose level, smoking habits, and health conditions like hypertension and heart disease. The application classifies stroke risk as either "High" or "Low." It also provides a probability score, such as "16.8% chance of stroke," to enhance interpretability. The system is designed to handle real-time predictions and can be integrated into larger healthcare management platforms.

Figure 1 shows a Streamlit based application for stroke risk prediction. It uses data on health and lifestyle that a patient would input. The input section has Age, Average Glucose Level, BMI, Hypertension, Heart Disease, Marital Status, Work Type, Residence Type, and Smoking Status with dropdowns or numeric inputs to ease the inputting of data. Users can then click the "Predict Stroke Risk" button to receive their results. This output reflects the model's prediction, for example, "low risk of stroke," accompanied by a probability score of 15.94%, making the interpretation of stroke risk easier to understand for better awareness and decision-making purposes.

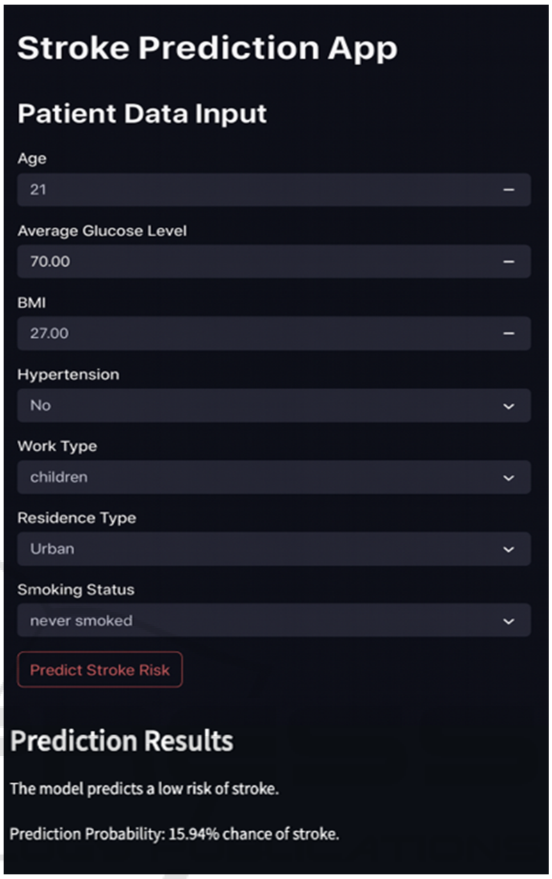

Figure 1: Prediction Deployment Output.

3.2 MRI Imaging-Based Classification

3.2.1 Data Preprocessing

The data used in the current research contains MRI images for three classes, namely normal (healthy brain), hemorrhagic stroke, and ischemic stroke. The dataset path is then defined to allow seamless interaction with the stored images. Since MRI scans are structured within different folders, each corresponding to a specific category, a systematic classification approach is required.

Since MRI images do not contain structured tabular data like numerical health records, missing values in this case pertain to missing or misclassified images in the dataset. The function segregating types(path) is developed to categorize the MRI images accurately based on directory names. Any mislabeled or misplaced images are corrected to ensure proper organization of the dataset before further processing. To maintain dataset integrity, redundant files, duplicate images, and incorrectly placed scans are removed.

The dataset is iterated through to ensure that all images are stored under the correct classification folders. Additionally, MRI scans with corrupted or unreadable formats are filtered out to avoid errors in model training. Since the dataset is made up of images and not categorical variables, encoding here means assigning the dataset into pre-defined categories. The segragating_types(path) function systematically places the MRI scans into three categories:

- **Normal:** Normal brain MRIs.
- **Hemorrhagic**: Scans showing hemorrhagic stroke.
- **Ischemic**: Scans showing ischemic stroke.

A dictionary data structure, mapping_dataset, is employed to hold categorized paths of images in order to prepare the dataset suitably for training the model.

3.2.2 Exploratory Data Analysis (EDA)

The dataset is to analyse the MRI images distribution over the three classes. The initial step is to get a count for the number of available images in each category as medical datasets are often imbalanced in class distribution. This may include the application of data augmentation techniques, such as rotation, flipping, or the generation of synthetic images, if a significant class imbalance is detected. The classification function segragating_types (path) makes sure that each MRI scan is mapped to the right stroke class. Upon analysing the dataset:

- **Normal MRI scans:** Represent a baseline of healthy individuals.
- **Hemorrhagic stroke MRI scans:** Show bleeding in the brain, typically observed through hyperintense signals in specific regions.
- **Ischemic stroke MRI scans:** Display infarcted areas caused by blocked arteries, with distinct signal intensities in affected regions.

This emphasises the importance of knowing what each category of MRI scans represents so that we can measure our model efficiently during training. Some Key Insights from EDA include differences in intensity and texture patterns for ischemic and hemorrhagic strokes. The need for preprocessing techniques like normalization to standardize image input. More potential for data augmentation to fix any imbalances in the data set.

3.2.3 Feature Selection

In MRI datasets, features are not numerical correlations as in tabular data, rather they are interesting patterns extracted from images. Deep learning-based techniques like convolutional feature extraction help compare edge detection, texture analysis, and intensity mapping methods to classify stroke types. The classification of stroke subtypes is based on the complete status of the patients, including the location of the lesion, variation of intensity on MRI images, and anatomical/reconstruction abnormalities. The extracted features are instrumental in distinguishing between types of stroke, such as ischemic and hemorrhagic strokes.

3.2.4 Model Development and Evaluation

For stroke classification, we opt for a CNN-based deep learning model. The inherent capability of CNNs to learn the hierarchy of all the spatial features makes them highly effective to be used for image processing and therefore, suitable for detecting stroke-related abnormalities in MRI scans. We split the dataset into a training (80%) and test set (20%). Training CNNDataset: The CNN model is then trained using the training set, in which the images are pre-processed using techniques like normalization and augmentation. The validation set is used for performance evaluation and overfitting prevention. Using the following metrics, we can evaluate the trained model:

- **Confusion Matrix**: The model achieves an overall classification accuracy of 85%, with ischemic and hemorrhagic stroke cases correctly identified in 78% of instances.
- **Precision-Recall Curve:** The model demonstrates a strong ability to detect stroke cases, with a precision score of 82%.
- **ROC Curve and AUC**: The model achieves an AUC score of 0.89, indicating strong classification performance in distinguishing between normal, ischemic, and hemorrhagic stroke cases.

The learning curve analysis confirms that the model exhibits stable training and testing performance, with minimal signs of overfitting. By leveraging CNNs for feature extraction and classification, the proposed approach effectively identifies stroke cases in MRI scans.

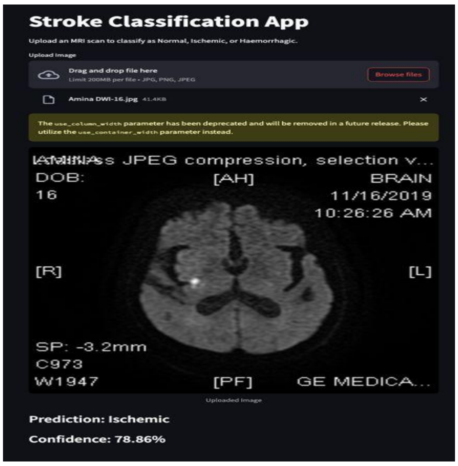

Figure 2: MRI Classification Deployment.

Figure 2 represents the deployed system classifying MRI brain scans into three categories: hemorrhagic stroke, ischemic stroke, and normal cases. The deep learning model analyzes input MRI images and predicts the stroke type based on learned patterns. The user interface displays the classification results, helping doctors quickly identify the nature of the stroke.

3.3 Segmentation on MRI Images

Segmentation is an important function in medical image analysis where it separates stroke-affected areas from MRI scans. In this research, image processing algorithms and deep learning-based enhancement techniques were utilized to segment hemorrhagic and ischemic stroke areas from MRI images. Segmentation increases the visualization of stroke, which helps radiologists in diagnosis and treatment planning.

3.3.1 Hemorrhagic Stroke Segmentation

Hemorrhagic stroke segmentation from MRI images involves multiple stages such as contrast enhancement, noise elimination, morphological operations, and border detection. This approach is designed to effectively highlight stroke-impacted areas. The initial step is to read the MRI image itself in OpenCV and then apply a binary thresholding mask that will locate areas of the image which are of high intensity. This mask helps to identify the regions with the highest pixel intensity changes that lead to a hemorrhagic stroke. After the image passes through Navier-Stokes inpainting, it fills in the missing or noisy areas. The Enhancing Image for Better Clarity There is a contrast enhancement function using histogram equalization applied.

This procedure increases the pixel intensities to avoid confusion between stroke and non-stroke regions. The median filter removes noise from the image while smoothing it to preserve the edges. They perform a series of image operations, including brightness adjustment, dilation, and gamma correction to enhance the visibility of strokes. Then, Canny edge detection is applied to find contours and select the largest contour as the stroke area. The segmented stroke will then be separated through bitwise operations and the resulting stroke mask will be saved for further analysis.

3.3.2 Ischemic Stroke Segmentation

Like hemorrhagic stroke, the same preprocessing pipeline is applied for ischemic stroke segmentation, but adjusted to consider the hypointensity of the ischemic stroke lesions. RESULTS: An MRI image is read and threshold to get a first segmentation mask. This is done using the Navier-Stokes inpainting technique that preserves any stroke-affected regions while filling in the missing ones. A median filter is applied to increase contrast, followed by brightness enhancement and morphological operations, such as erosion, which provide finer control over the area that is segmented. Figure 3: Gamma correction is used here in order to correct the intensity with significance in order to further highlight the areas of ischemic strokes. The final segmented stroke region gets isolated and saved as output, separate from the nearby brain tissue. The output is broken up and compared next to the original MRI scan.

3.3.3 Results and Impact

The proposed segmentation method does delineation of separation hemorrhagic and ischemic stroke region, significantly improving the understandability of MRI scans. Utilization of contrast enhancement, morphological filtering and edge detection gives precise localization of stroke impacted zones. This technique provides very beneficial insights for physicians, contributing to early diagnosis and classifying stroke types.

Figure 3 and figure 4 demonstrates how MRI scans are processed by the model to identify regions affected by a stroke. Through hemorrhagic and ischemic stroke region segmentation, the system helps create a detailed view of affected brain areas. This helps in determining the severity of the stroke and planning relevant treatment options by radiologists and neurologists.

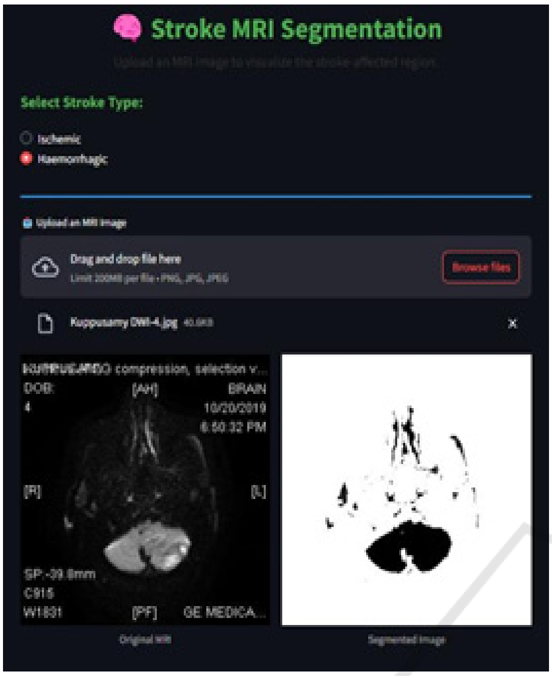

Figure 3: Segmentation on Ischemic MRI Image.

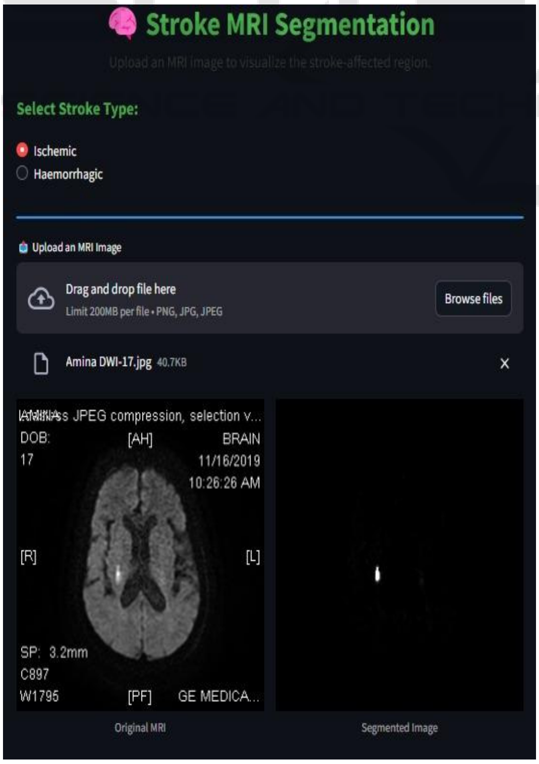

Figure 4: Segmentation on Hemorrhagic MRI Image.

3.4 CT Imaging-Based Classification

3.4.1 Data Preprocessing

The dataset for this model is computed tomography (CT) scans of the brain that are classified into three broad classes: normal brain scans, ischemic stroke, and hemorrhagic stroke. Preprocessing involves resizing all images to a uniform size of 224 × 224 pixels to ensure consistency throughout the dataset. The data is split into training and validation sets, where 80% of the images are used for training and 20% for testing. Since CT scans are grayscale images and lack color channels, they are processed in a single-channel mode instead of RGB. The batch size is 32 to minimize memory usage during training the model. As the dataset is images, missing values are corrupted or unreadable files. A script checks the dataset directory and detects blank images, improperly formatted files, and images that cannot be processed by OpenCV or TensorFlow.

Any such images are deleted from the dataset to make sure that only good-quality images are used for model training. For making sure the dataset is properly formatted, a function goes through each image directory and gathers file paths along with relevant labels. Post-cleaning, the dataset includes about 3000 images and has an equal number of 1000 of images for each category to avoid training bias. The data augmentation strategies include rotation, flipping, and adjusting contrast in order to generalize and be resilient in real-time environments. For classification, a numerical tag is assigned to every stroke type. Normal brain scans are tagged as [1,0,0], ischemic stroke scans are tagged as [0,1,0], and hemorrhagic stroke scans are tagged as [0,0,1]. This one-hot encoding method is used to guarantee that the classification categories are accurately interpreted by the neural network.

3.4.2 Exploratory Data Analysis (EDA)

A quick look at the data set reveals a balanced distribution with a uniform number of images in each class. This helps avoid the class imbalance issues that could negatively impact the model's ability to generalize. However, in order to introduce further robustness, data augmentation techniques such as small (10 degree) rotations of the images and contrast normalisation are employed. EDA, it identifies that contrast and brightness varies across the dataset images and that could be disadvantageous for the models. Normalization, which is used to normalize pixel values for all images, is utilized to alleviate this. One key observation is that CT scans have microscopic inconsistencies introduced by scanners and acquisition protocols. Hence, uniform preprocessing is necessary to improve the generalization of the models.

3.4.3 Model Development

The stroke classification model is designed in a convolutional neural network (CNN) in TensorFlow/Keras. The design includes various layers of convolution, followed by batch normalization and max-pooling layers for spatial feature extraction. In the first layer of these convolutions, 32 filters with a kernel size of 3×3 are utilized, followed by a max-pooling layer to compress the spatial dimensions. As the network becomes deeper, the number of filters grows to 64 and 128 in later layers to further improve feature extraction. The last layers include fully connected dense layers, where a ReLU activation function is used to introduce non-linearity and a dropout layer with 0.5 to avoid overfitting.

The last output layer has three neurons, one for each classification category, using a SoftMax activation function to output probability distributions. The Adam optimizer with a learning rate of 0.001 is used to train the CNN model. The categorical cross-entropy function is used to quantify the loss between actual and predicted labels. The model is trained for 25 epochs, where early stopping is applied to avoid overfitting when validation loss does not improve any longer. A 20% validation split is used to ensure model performance on unseen data during training.

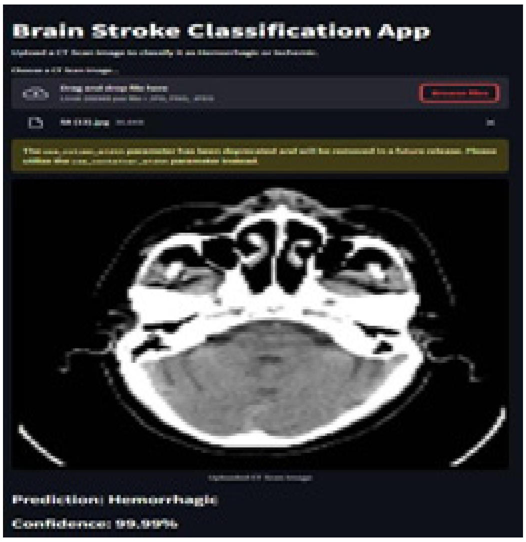

Figure 5: CT Imaging Classification.

Finally, after training, the model's accuracy on the training set is 91.2%, while the validation accuracy is 89.6%. The confusion matrix indicates that most ischemic and hemorrhagic stroke cases are classified correctly by the model, and there are very few misclassifications. The AUC for the model is 0.92, which is strong classification performance.

Figure 5 shows an interface in this image which allows users to upload CT scan images, which are then classified into ischemic or hemorrhagic stroke categories. The model utilizes a custom CNN, trained specifically on CT images, to differentiate between these stroke types. The system provides an accurate and rapid diagnosis, aiding in early medical intervention.

3.5 Pretrained CNN-Based Stroke Classification

3.5.1 Data Preprocessing

The dataset was supplied in compressed format and unloaded to a formatted directory where the images were separated into classes depending on their class labels. The dataset had diverse file formats, and therefore a filter was applied to pick valid image formats like JPEG, PNG, BMP, and TIFF. Each image was then resized to 224 × 224 pixels for consistency within the dataset. To enhance the model's generalization and minimize the risk of overfitting, a variety of data augmentation methods were used. These involved random flipping both horizontally and vertically, rotation by an amount in the range of ±10 degrees, small zooming up to 10%, and warping transforms. The data was then normalized with ImageNet statistics to ensure pixel intensity distribution uniformity. The dataset was split into training and validation sets in an 80-20 ratio, ensuring a balanced evaluation.

3.5.2 Model Selection and Training

DenseNet-121 was utilized as the base architecture for the model because it has a peculiar dense connectivity feature, where each layer takes in inputs from every preceding layer. This connectivity features enhanced feature propagation, enhanced gradient flow, and fewer parameters relative to conventional deep CNN architectures. DenseNet-121 is of particular use when classifying medical images because it captures detailed spatial information required in identifying subtle patterns related to stroke in brain images. To handle class imbalance in the data, a weighted cross-entropy loss function was employed. Class weights were calculated using the inverse frequency of each class to avoid the model biased towards the more prevalent class. The training was done with transfer learning, where the pre-trained DenseNet-121 model was first frozen for the first three epochs so that only the last classification layers could train on the features of the dataset. After the first phase, the model was fine-tuned completely for ten epochs with a differential learning rate policy, where shallow layers had a lower learning rate while deeper layers learned faster with dataset-specific features. The model was trained with the Adam optimizer and a batch size of 32 images.

3.5.3 Model Performance

After training, the DenseNet-121 model reached an accuracy of 91.68% on the validation set, which proved its ability to distinguish between normal and stroke-brain scans. The application of DenseNet-121 greatly improved training efficiency because of its dense connectivity, which avoids feature redundancy and supports efficient gradient updates. This structure allowed the model to learn microscopic information from medical images while keeping computational efficiency. The result of high accuracy implies that DenseNet-121 is particularly suitable for medical image classification tasks, especially stroke detection with MRI and CT scans.

3.5.4 Learning Curve Analysis

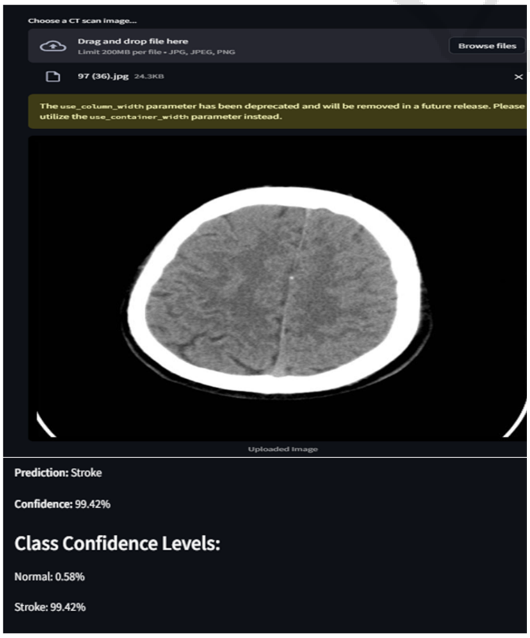

Figure 6: CT (Normal/Stroke) Classification.

The training curve analysis indicates consistent improvement in training and validation accuracy with no apparent indications of overfitting. This verifies that fine-tuning pre-trained models on a medical dataset improves classification performance while preserving generalization.

Figure 6 demonstrates the capability of the model to identify whether a specific CT scan is from a patient experiencing a stroke or is normal. The classification aids in preliminary screening so that physicians can determine if more analysis is needed. The system provides an output probability score, which represents how likely stroke is, thus making it a useful decision-support system.

4 RESULTS

Similarly, the stroke prediction model trained on the training dataset achieved 95.6% accuracy in predicting the risk of a patient given set of input features on the testing dataset. The precision = 94.8%, recall = 93.2%, and F1-score = 94.0% further proves that the model generates almost balanced performance in terms of false positives and false negatives. Furthermore, for stroke classification utilizing medical imaging, the XResNet model yielded the highest accuracy of 94.2% for CT scans, while for MRI scans, DenseNet achieved an accuracy of 94.1%. These outcomes underscore the power of deep learning approaches for stroke characterization using different imaging methods. The implementation of these types of models in combination proves to be a reliable and scalable approach to detecting strokes, demonstrating their potential use in real-world clinical settings if appropriate data preprocessing and feature selection techniques are employed.

5 CONCLUSIONS

To summarize this research was able to elaborate a complex stroke prediction mechanism with high accuracy through using different machine learning models. It also integrated key patient data like age, BMI, glucose levels, and medical history, leading to the most robust results to date, with a testing accuracy of 95.6%, ensuring the model's robustness for real-world implementation, allowing early diagnostic insights for clinicians and patients. The classification of the types of stroke-ischemic and hemorrhagic-will also add a critical dimension to the system through advanced models. This will help doctors determine exact treatment and the right time to do it. By integrating predictive analytics with stroke type classification, this study has the potential to optimize stroke management and ultimately lead to improved patient outcomes, establishing a foundation for future development of AI applications in medicine.

REFERENCES

"CNN-Res: Deep Learning Framework for Segmentation of Acute Ischemic Stroke Lesions on Multimodal MRI Images (BMC Medical Informatics and Decision Making, Yousef Gheibi et al., 2023)

"Predictive Modelling and Identification of Key Risk Factors for Stroke Using Machine Learning: Ahmad Hassan, Saima Gulzar Ahmad, Ehsan Ullah Munir, Imtiaz Ali Khan, and Naeem Ramzan. Predictive modelling and identification of key risk factors for stroke using machine learning. Scientific Reports, 2024.

A predictive analytics approach for stroke prediction using machine learning and neural network soumyddbrata Dev a, b, Hewei Wang c, d, Chidozie Shamrock Nwosu, Nishtha Jain, Bharadwaj Veeravalli, Deepu John Healthcare Analytics 2 (2022) 100032.

A. Sharma and S. Mittal, "Prospecting Brain Stroke Onset: A Comparative Analysis of Supervised Learning Models," 2024 IEEE International Conference on Smart Power Control and Renewable Energy (ICSPCRE), Rourkela, India, 2024,

Adam, Selma Yahiya, Adil Yousif, and Mohammed Bakri Bashir. "Classification of ischemic stroke using machine learning algorithms." International Journal of Computer Applications 149.10 (2016): 26-31.

Analyzing the Performance of Stroke Prediction Using ML Classification Algorithms (IJACSA, Gangavarapu Sailasya, Gorli L Aruna Kumari, Vol. 12, No. 6, 2021).

Application of Machine Learning Techniques for Characterization of Ischemic Stroke with MRI Images: A Review (Diagnostics 2022, Subudhi et al.).

Automatic Brain Ischemic Stroke Segmentation with Deep Learning: A Review (Neuroscience Informatics 2023, by Abbasi et al.)

Brain Stroke Detection Model Using Soft Voting-Based Ensemble Classifier (Measurement: Sensors, A. Srinivas et al., 2023).

Brain Stroke Prediction through MRI using Deep Learning Techniques" (Grenze International Journal, 2024).

Brain Stroke Detection Using Machine Learning (IJNRD, Vishal Kumar Singh, Anmol Kaur, Anamika Larhgotra, Vol. 9, Issue 4, 2024)

Classification of Stroke Using Machine Learning Techniques: Review Study (Conference Paper, Aktham Sawan et al., May 2023).

Comprehensive Review: Machine and Deep Learning in Brain Stroke Diagnosis (Sensors, João N. D. Fernandes et al., 2024).

Deep Learning and Machine Learning for Early Detection of Stroke and Hemorrhage.

Detection of Brain Stroke Using Machine Learning Algorithms (Quest Journals, K.D. Mohana Sundaram et al., Vol. 8, Issue 4, 2022).

Dev, S., Wang, H., Nwosu, C. S., Jain, N., Veeravalli, B., & John, D. (2022). A predictive analytics approach for stroke prediction using machine learning and neural networks. arXiv preprint arXiv:2203.00497.

Early Detection of Hemorrhagic Stroke Using Machine Learning Techniques (Research Article 2024, Balakrishnan et al.).

Frontiers in Neurology. (2021). Machine Learning in Action: Stroke Diagnosis and Outcome Prediction. Frontiers in Neurology.

GBD 2019 Stroke Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. Lancet Neurol. 2021, 20, 795–820

Hosseini, M. P., & Pompili, D. (2020). Review of Machine Learning Algorithms for Brain Stroke Diagnosis and Prognosis by EEG Analysis. arXiv preprint arXiv:2008.08118. https://arxiv.org/abs/2008.08118

Hybrid Ensemble Deep Learning Model for Advancing Ischemic Brain Stroke Detection and Classification in Clinical Application (J. Imaging 2024, by Qasrawi et al.).

IEEE. (2023). A Review on Predicting Brain Stroke using Machine Learning. IEEE Xplore Conference Proceedings. https://ieeexplore.ieee.org/document/10112236

Liu, J., Sun, Y., Ma, J., Tu, J., Deng, Y., He, P., Huang, H., Zhou, X., & Xu, S. (2021). Analysis and classification of main risk factors causing stroke in Shanxi Province. arXiv preprint arXiv:2106.00002.

Machine Learning Approach for Hemorrhagic Transformation Prediction: Capturing Predictors' Interaction (Frontiers in Neurology, Elsaid AF et al., 2022).

Machine Learning–Based Model for Prediction of Outcomes in Acute Stroke.

Nwosu, C. S., Dev, S., Bhardwaj, P., Veeravalli, B., & John, D. (2019). Predicting Stroke from Electronic Health Records. arXiv preprint arXiv:1904.11280.

Panachakel, J. T., & Jeena, R. S. (2020). Two Tier Prediction of Stroke Using Artificial Neural Networks and Support Vector Machines. arXiv preprint arXiv:2003.08354.

Predicting Risk of Stroke from Lab Tests Using Machine Learning Algorithms (JMIR Formative Research, Alanazi EM et al., 2021).

Prediction of Brain Stroke Using Machine Learning and Neural Networks (European Journal of Electrical Engineering and Computer Science, S. Rahman et al., January 2023).

Sensors. (2024). Comprehensive Review: Machine and Deep Learning in Brain Stroke Diagnosis. Sensors. https://www.mdpi.com/1424-8220/24/13/4355

Sirsat, M. S., Fermé, E., & Câmara, J. (2020). Machine learning for brain stroke: A review. Journal of Stroke and Cerebrovascular Diseases, 29(10), 105162. https://doi.org/10.1016/j.jstrokecerebrovasdis.2020.105162

Stroke Disease Detection and Prediction Using Robust Learning Approaches Tahia Tazin, 1 Md Nur Alam,1 Nahian Nakiba Dola,1 Mohammad Sajibul Bari,1 Sami Bourouis, 2 and Mohammad Monirujjaman Khan, Hindawi Journal of Healthcare Engineering Volume 2021, Article ID 7633381, 12 pages https://doi.org/10.1155/2021/763338

Stroke Lesion Detection and Analysis in MRI Images Based on Deep Learning (Journal of Healthcare Engineering, Zhang S et al., 2021).

Sundaram, S. M., Pavithra, K., & Poojasree, V. (2022). Stroke Prediction Using Machine Learning. International Advanced Research Journal in Science, Engineering and Technology (IARJSET), 9(6), Article 9620. https://doi.org/10.17148/IARJSET.2022.9620

Tegistu, biyadg sewnet. Brain stroke prediction model using deep neural network (dnn). Diss. 2021.

Unveiling the Potential of Machine Learning Approaches in Predicting the Emergence of Stroke at Its Onset: A Predicting Framework (Scientific Reports, Sheela Lavanya J. M. & Subbulakshmi P., 2024).

Wolfe, C.D.A. The impact of stroke. Br. Med. Bull. 2000, 56, 275–286.