Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Brain Stroke Detection and Classification System: A Hybrid Approach using Deep Learning Techniques

Avyukth Inna[1], Selva kumar S[2] Ayman Khan[3], Brijesh S G[4], Gurrala Naga, Pragnathmik[5]
[2]*Associate Professor, Computer Science & Engineering, B.M.S. College of Engineering, Bangalore, Karnataka ,*
Orchid id-0000-0002-3342-7161, selva.cse@bmsce.ac.in
[1,3,4,5]*Department of CSE, B.M.S College of Engineering, Bengaluru*

ARTICLE INFO

Received: 30 Dec 2024

Revised: 19 Feb 2025

Accepted: 27 Feb 2025

ABSTRACT

Brain stroke is a major health concern around the globe, which causes significant disability and loss of life. Effective treatment can be ensured only with accurate and timely diagnosis. However, traditional methods for stroke diagnosis heavily rely on manual evaluation of patient symptoms and CT scans, which prove to be time-consuming and prone to errors. Our major objective of this study is to develop a reliable stroke detection and classification system that can automate the diagnosis process using deep learning techniques. The proposed pipeline integrates a MLP model analysis textual patient health records and a CNN model for image based classification of CT scans. The study aims to provide a fast, reliable and systematic tool to aid medical professionals in detecting stroke accurately and removing any hindrance in treatment. In addition to the proposed methods, other models including ResNet50 and VGG for stroke classification and Random Forest for text analysis were also explored. The MLP model achieved accuracy of 94.67% for stroke detection based on patient data while the CNN displayed superior performance with accuracy of 98.6%. Relative analysis with accuracy, precision-recall and F1 score confirmed CNN beat pre-trained models like VGG and ResNet50, while the MLP model outperformed the Random Forest Classifier.

Keywords: Brain Stroke, CT Scans, Deep Learning, MLP, CNN, SMOTE

INTRODUCTION

Brain Stroke is one of the leading causes of death worldwide. Strokes occur when blood circulation in brain sections is blocked, leading to the cell death due to the increase in lack of oxygen [1]. There are two types of stroke: ischemic and hemorrhagic [2]. Ischemic stroke occurs due to a clot in a blood vessel supplying a tissue or region. On the other hand, hemorrhagic strokes occur when a blood vessel bursts which usually causes internal bleeding. The WHO states that in lower income countries and middle income countries, there are more than 11 million strokes every year and this number continues to escalate with passing years due to demographic changes and lifestyle shifts. This leads to almost 4 million deaths every year and nearly 35% of these survivors are left disabled. The most common symptoms include weakness, speech difficulties, and intense migraines. Interventions taken immediately after an event can greatly increase the chances of recovering, hence why getting diagnosed and treated is vital. Treating procedures for the two different strokes vary extensively, and hence why misdiagnosis or delays in determining the detection and categorization of strokes may result in harmful treatments that worsen brain damage or death. Currently, diagnosis especially in resource-limited settings depends on standard manual interpretations.

Stroke detection remains to be a time sensitive challenge due to symptom complexities and necessity of rapid intervention. Moreover, early-stage strokes may have benign symptoms, making it difficult for physicians to provide immediate diagnosis. Recent advances in machine learning have demonstrated potential medical diagnosis. Image processing techniques based on Deep Learning have achieved near-human accuracy in detecting medical abnormalities across multiple healthcare domains, making it confidently capable of stroke detection and classification purposes [3]. Likewise, structured patient records that have been a vital factor towards stroke detection can be effectively analysed leveraging various ML and DL algorithms to inturn assess stroke risk.

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

The confluence of the need for a fast and accurate stroke diagnosis system and the growing capabilities of AI enables the development of an effective stroke detection system, which can assist local clinics reducing the need for expensive tests. Existing automated systems for stroke detection are limited in their approach, either only focusing on patient attributes or medical imaging such as CT or MRI scans. These only offer limited scope in the level of accuracy that can be achieved diminishing the reliability of the system. To address this limitation, the paper presents a hybrid model of deep learning where both image and attribute data are part of the pipeline in detecting Stroke. Specifically a MLP mode is utilized for prediction based on patient attributes and a CNN is used to classify potential stroke subtypes. The system is designed to be rapid and accurate for clinicians to employ and use.

OBJECTIVES

The paper's main goal was to develop a large system that combines patient records with unstructured CT Scan images. This approach aims to provide a complete view of stroke diagnosis, using various data types to improve accuracy.

The system aims to cut down the dependence on laborious image interpretation which is often a long and error prone process. By using automated systems, this study aims to get faster and dependable results which can be vital in time sensitive situations. Inaccurate diagnosis or delays in identifying stroke subtype can end in unsuitable treatments or procedures increasing the risk for the patients.

Prompt detection and classification of stroke is paramount to improving patient health. Creating a system for rapid detection and classification decreases the risk of inaccurate diagnosis. Moreover, stroke diagnosis presents a unique challenge in regions with low income or limited medical resources, this project aims to create a low cost and effective system to narrow the disparity in stroke care, reducing the burden on the medical infrastructure.

LITERATURE REVIEW

A detailed study of the literature and methods of stroke detection and classification was performed in an attempt to learn more before the proposed system was designed. A number of studies have aimed at the various approaches of deep learning and machine learning, which have been demonstrated to produce favorable outcomes in stroke detection for clinical images.

One of the newest techniques uses Convolutional Neural Networks (CNN) to identify stroke from images such as CT and MRI scans. Gahiwad et al. [4] used a CNN model to identify brain stroke from CT scan images with an accuracy of 90%. The dataset contained 2,551 images. The study acknowledges the small size of the dataset as a limitation. On the other hand, Saleem et al. [10] proposed a hybrid approach to using CNN in combination with machine learning classifiers like Support Vector Machine (SVM), Naïve Bayers and Random Forest (RF). This approach achieved an accuracy of 98.43% using the dataset which consisted of 2,500 images and 10,000 patient records, which show the potential of the model in stroke detection from MRI scans. Other studies like the one conducted by Mushtaq and Saini [6] also utilized CNN with techniques like data augmentation to improve the model's accuracy and reduce overfitting.

There has been recent emphasis on hybrid approaches, which combine CNN with other machine learning techniques to improve the performance of the model in stroke detection. For example, Gaidhani et al. [7] employed a hybrid CNN with Long Short-Term Memory (LSTM) models in an attempt to classify stroke types following feature extraction using CNN. Their model achieved an accuracy of 98.42% on CT scan data sourced from Kaggle. The AUC score achieved was 0.99 which demonstrates the effectiveness of the hybrid approach. Furthermore, Mohammed et al. [8] put forth using transfer learning techniques alongside CNN for detecting brain stroke. In a similar approach Shinde et al [9] combined ensemble classifiers with deep learning methods which showed promising results.

The feature extraction and data augmentation process are critical in improving the performance of models used for stroke detection. Some studies have used conventional image processing methods while others like Gahiwad et al. [4] used data augmentation to reduce overfitting. Similarly Rahman et al. [10] studied the use of Particle Swarm Optimization (PSO) to optimize the feature extraction from medical images which resulted in an accuracy of 90% in detecting many kinds of hemorrhagic strokes from CT scan data. The research highlights how data augmentation and optimisation can enhance the accuracy of the models significantly.

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Alongside deep learning models, many machine learning classifiers have been assessed for stroke prediction. One study by Victor et al. [11] used an ensemble approach by combining LightGBM, Random Forest and XGBoost. Random Forest achieved maximum accuracy of 99% reflecting its potential in successful stroke prediction. In the research work performed by Srinivas et al. [1], soft-voting classifier with a blend of classifiers like Random Forest, Decision Tree and SVM achieved an exceptionally good result of 96.88% accuracy and 0.97 F1 score.

Notwithstanding these advancements, there are still challenges in stroke detection, mainly concerning the size of the dataset, overfitting and overall computational cost. Several studies have noted the limitation of a small dataset in the development process of models for stroke detection and classification. For example, Saleem et al. [5] and Rahman et al. [6] identified that smaller dataset could affect the model's ability to generalize its predictions which may lead to overfitting when there are other complex architectures involved. Moveover, a study by Alruily et al. [12] proposed a Particle Swarm Optimization (PSO) in combination with Federated Learning model to improve the accuracy while ensuring the privacy of the patient's data.

In conclusion, the literature highlights the effectiveness of machine learning and deep learning methods for detecting brain stroke. CNN-based approaches dominate the field and hybrid models that integrate machine learning techniques for classification, feature optimization and data augmentation are emerging as strong candidates for developing new models with better accuracy and robustness. Nonetheless, these challenges must be addressed to improve the systems for brain stroke detection and classification.

METHODOLOGY

Once the literature survey was finished, the methodology of this study was divided into two primary sections: Stroke Detection and Stroke Classification. Each of the subsections provide a comprehensive interpretation of various steps performed to eventually attain the presented hybrid system.

Stroke Detection

Data Collection and Analysis

We first meticulously validated the collected data before model training began. This step required looking at the feature distribution, missing value treatment, class imbalance treatment, and outlier detection that was likely to cause blunders.

The text datasets used for analysis were acquired from Harvard Dataverse, which has more than 43000 entries along with 12 different features. Each entry or row was a single patient's health record which comprised both clinical and demographic data that was most important for stroke prediction. The dataset contained categorical features for analysis such as gender and marital repute, work type, residence type, and smoking status, as well as quantitative variables such as age, mean glucose level, and Body Mass Index. In addition, critical medical variables including an account of hypertension and coronary heart disease were included to evaluate the undeclared impact of stroke [13]. To understand better how the stroke risk negatively and positively interacted with the numerical variables, a correlation matrix was prepared and for every numerical feature, the Pearson correlation coefficient was calculated. Linearity ratio of two variables is defined by the Pearson correlation coefficient and is presented with r cost between -1 up to 1, in which 1 signifies a strong positive correlation, -1 signifies a strong negative correlation, and 0 indicates no correlation. The correlation matrix illustrated the ways in which factors like a person's age, average glucose level, and BMI contribute to the risk of having a stroke.

Age carried the strongest positive correlation with stroke, though still weak, highlighting that additional categorical variables may have a stronger predictive value. Glucose levels and BMI, when studied separately, did not influence stroke directly. In order to further understand the role of demographic and medical factors on the prevalence of stroke, gender and hypertension were also analyzed categorically in relation to stroke. Among the female participants, 1.08% were found to have had a stroke which is slightly lower than the male participants' rate of 1.32%. This showed the relative equilibrium of stroke occurrence between the two genders while confirming the rationale behind the scientific literature which asserts strokes are more frequent among men as a result of their unhealthy lifestyle combined with hormonal and cardiovascular risk factors.

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Hypertension was regarded as one of the most important risk factors of having a stroke and the dataset confirms this assertion. The rate of stroke occurrence in people without hypertension was only 0.96%. But, those who have been diagnosed with hypertension had a staggering rate of 4.16%. This substantial difference highlights how crucial hypertension treatment is to preventing strokes. This emphasizes the need for early detection, lifestyle changes, and treatment for the medically diagnosed high blood pressure patients.

In conclusion, to verify data completeness, boxplots and histograms were created to check for outlier values. Outliers are data points that dramatically differ from other data points and can be highly problematic during the training of machine learning models. In fact, they can result in misleading predictions due to introducing bias and skewing the statistical distributions. Interquartile Range method was used with a for loop in Python, making these outliers disappear from the dataset. This method was executed until there were no more extreme values providing a more reliable set of data.

Data Preprocessing

In data preprocessing phase for MLP text model, several steps were performed to ensure the dataset was well ready for training. Initially, the dataset had a heavy-class imbalance, with stroke cases to non-stroke cases. To address this, we applied Synthetic Minority Oversampling Technique (SMOTE) [14]. By creating synthetic samples from the minority class using interpolation, SMOTE reduces data imbalance and consequently enhances the learning process of models hung up in predicting the majority class. After balancing the dataset, feature scaling was carried out using the StandardScaler from Scikit-learn. This step finally normalized the data, having mean 0 and the standard deviation equal to 1 for each feature. Thus, the neural networks' training time can be reduced by normalization by ensuring that all features contribute equally to the model, making it an extremely crucial step [13].

To refine the dataset, we used Recursive Feature Elimination (RFE) with an SVM estimator. RFE works by removing less important features one by one until only the important ones are left in the dataset. Furthermore, the dataset was divided into the training and the test group, at an 80 to 20 percentage split. This way, the model learned from most of the data but still had a separate part to evaluate how well it performed. The target variable, "stroke," was coded with 1 for when a stroke happened and 0 for no stroke risk.

In summary, the following sequence was adopted for preprocessing the text data: SMOTE for oversampling, StandardScaler for normalization, RFE for feature selection, data splitting, and finally target encoding. These processes assisted in preparing the dataset to build a robust and accurate MLP model.

Proposed Model Architecture

The model Multi-Layer Perceptron (MLP), is a variation of a feedforward neural network and it is characterized by containing several stacked layers of nodes that include: an input layer, one or multiple hidden layers, and output layer. Each node in a layer is connected to all other nodes in the ensuing layer, which have weighted connections that can be modified during training to minimize prediction error [15].

In the MLP model architecture, there is an input layer of 10 neurons, which corresponds to the 10 selected features that were chosen from the dataset. These features were selected using Recursive Feature Elimination (RFE) in order to make sure the most relevant features for training are used. The hidden layers of the MLP consist of three layers with 124, 64, and 32 neurons, respectively. All hidden layers are equipped with ReLu (Rectified Linear Unit) activation, which allows non-linearity to be added to the model and improves the convergence speed of stochastic gradient descent to the model when compared to sigmoid or tanh functions [10]. The mathematical formula of the ReLu function is given in Equation (1).

$$f_{\mathrm{ReLU}} = max\,(0, x) \qquad (1)$$

To prevent overfitting, dropout layers with a dropout of 0.3 were added after two hidden layers where a set fraction of input units were randomly set to zero during training. In the MLP model, the output layer has a single neuron and uses sigmoid activation function that provides outputs between 0 and 1. This is the stroke probability; that is, values close to 1 have a greater chance of causing a stroke. A few of the weight parameters are updated by using AdamW optimizer on the model after training, which is weight decay regularized modified Adam optimizer. This is done to

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

further prevent overfitting by using larger weight penalization while training. The loss function used was BCEWithLogitsLoss, which combines sigmoid activation and binary cross entropy loss to provide stability and efficiency. The model was trained for over 150 epochs with learning batch size 64. The initial learning rate was 0.001 and weight decay parameter was 1e-5 while training. While training, the performance of the model was monitored with a validation group to make sure that it was learning correctly and not overfitting the training data. The training hyperparameters are presented in Table 1.

Table 1. Hyperparameters used in the MLP model

Training Parameters					
Learning Rate	**Batch Size**	**Optimizer**	**Loss function**	**No. of Epochs**	**Metrics**
0.001	64	AdamW	BCEWithLogitsLoss	150	Accuracy

In conclusion, the MLP model used is a robust neural network architecture with the purpose of predicting the occurrence of brain stroke from textual health parameters. The dropout layers, weight decay regularization, and extensive hyperparameter tuning provided further accuracy and generalization, making it a valuable tool for early stroke detection.

Stroke Classification

Data Collection and Analysis

We collected over 2500 CT Scan images from Kaggle for the stroke classification phase and they were categorized into ischemic and hemorrhagic classes. These images formed the dataset for training the CNN model later capable of distinguishing between the two effectively. Before training the model, we made a thorough analysis of all images to ensure credibility. This included checking for the class balance as an imbalance dataset could lead to biased model performance. All images were carefully examined for corrupt or unreadable files, which were either repaired or removed to prevent inconsistencies during training.

The dataset nature was attained by conducting an initial analysis of the CT Scan images. A histogram of pixel intensity was drawn for a group of images with contrast, brightness, and noise. Image flattening into 1D arrays and binning of the pixel intensities into a total of about 50 bins were accomplished using python functions. This helped spot intensity patterns throughout the dataset. Besides, we built the mean image representation by stroke class to determine the most frequent stroke patterns, intensity gradients, and potential imaging differences. We achieved this by starting with the conversion of all the images to grayscale, down-sampling them to a consistent resolution of 128×128 pixels, and splitting them based on their respective labels.

A representative mean image was then built for ischemic and hemorrhagic stroke classes, by computing the average pixel intensities of all images per class. The mean images showed stroke patterns and intensity trends, enabling the model to learn informative stroke features.

To better learn the pixel-level variation in each stroke class, we performed statistical analysis. After preprocessing, the pixel intensities of all the images were flattened and binned based on their respective stroke labels. Mean pixel intensity was computed for each class to derive the average structural features of the stroke class in the form of a representative image. In parallel, the standard deviation was computed to measure the variation of the pixel intensities in the class. For structural feature analysis, edge detection was carried out on a collection of images. Edge detection is simple to apply for feature extraction in training the model because it encourages salient boundaries and inhibits noise. After standardization, an edge detection filter was applied to extract high-frequency content from the images with focus on boundaries and textural features.The filter does this by responding to pixel intensity gradients, high contrast transition areas such as boundaries between different densities of tissue or stroke lesions in brain scans.

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

To be used as a visual comparison of the impact of the transformation, the original grayscale image and the edge-detected image were presented side by side. The original image preserves the entire intensity distribution, but the processed image accentuates structural edges and areas of interest with high variance. The operation is particularly valuable in stroke classification application as it offers convenience in ischemic and hemorrhagic pattern discrimination, which are most often of different textural features.

Data Preprocessing

After proper inspection, all the images were preprocessed in a systematic manner to ensure consistency. This involved loading, resizing, and arranging the dataset in a manner that preserved clear differences between the two types of strokes. All the images were resized to 128×128 pixels to standardize input size and therefore, prevent differences in dimensions from affecting model training. Once all images were processed, they were stored as arrays with corresponding class labels. The dataset was split into training (80%), validation (10%), and testing (10%) sets.

Raw images have pixel values from 0 to 255, which can cause unstable weight updates. To address this, we applied min-max normalization and thus; scaled all pixel values to a normalized variety between 0 and 1. This transformation divides each pixel value by 255 and ensures that every feature contributes similarly to the learning process, further avoiding issues related to large gradients or exploding values.Since the class labels of the stroke class were originally in string format, they first needed to be converted into numeric values before being passed into the algorithm. Label encoding was used to accomplish this, where each unique stroke class was assigned a unique integer value. The translation allowed the categorical labels to be treated as numbers, which was sufficient for the classification layer of the model. This optimized dataset then served as a better input for the classification model, which became more efficient and reliable.

Proposed Model Architecture

Convolutional Neural Networks transformed the Computer Vision world by allowing machines to learn and derive hierarchical features from images without any human intervention. The biological visual pathway being emulated, CNNs apply stacks of convolutional filters to search through images, extracting spatial relations and characteristic features required for purposes of classification [16].

The Convolutional Neural Network design suggested is based on extracting CT-Scan image stroke features with a compromise among feature extraction, computation cost, and regularization for optimizing the performance. The first convolutional layer accepts preprocessed input of size 128×128 and made up of 8 filters with size 3×3.Each filter scans the input image and extracts features like textures, edges and patterns. The Rectified Linear Unit activation function (1) instigates non-linearity making sure the model is able to learn complex patterns, preventing problems like vanishing gradient and thus, speed up training. A L2 regularization term (0.01) is applied to this layer to prevent overfitting by penalizing large weight values.After the features are extracted using convolution, the model uses Max Pooling to further reduce and simplify the data without losing important information. The process is carried out by scanning small regions of the image and selecting the maximum values, thereby retaining the maximum patterns but losing secondary information. Max pooling reduces the overall image representation with this, making the model more efficient and less prone to overfitting. There is a single max pooling layer after each convolution layer in the model in order to maintain consistency of feature extraction.

As the network deepens, the model introduces convolutional layers with increasing filter sizes, each one with their own ReLu function strengthening the ability of the model to recognise complex patterns. To enhance model stability, batch normalization is added at major points within the architecture. The technique normalizes the inputs to standardize the activations in each layer by keeping the distribution of the data consistent throughout the training process. The second, third, and fourth convolutional layers have 32, 64 and 64 kernel filters respectively.

The model then transitions from convolutional to a fully connected one. The last flattening layer converts the spatially learned features into a one-dimensional vector, which is now ready to be classified. The first Dense layer adds 128 neurons for deep feature extraction, while an additional layer refines representations. In the last part of the model, a single neuron is specified to provide the final classification. This neuron gives a probability score between 0 and 1 via sigmoid activation function (2). This ensures the model produces an interpretable result, with closer values to 0 being ischemic and closer values to 1 being hemorrhagic. The sigmoid activation function equation is given in Equation 2.

$$f_{\text{SIGMOID}} = \frac{1}{1 + e^{-(x)}} \quad (2)$$

To ensure optimal learning and efficient weight updates, the model was compiled using Adam optimizer technique that dynamically adjusts the learning rates of each parameter [14]. This technique combines momentum and adaptive learning rate to accelerate convergence and avoid local minima. Binary cross entropy was chosen as the loss function for measuring the gap between the predicted probability and actual class label, ensuring that the model learns to make accurate predictions. It penalizes overconfident wrong predictions, helping the model improve decision-making over time. During training, the dataset was fed into the model in small batches. This technique optimized computational efficiency while maintaining stable updates to the model's internal parameters. The model trained over 20 epochs; with 32 images in each batch for every epoch. After each epoch, the performance was evaluated on a separate validation group, allowing for real-time tracking of progress and detecting potential overfitting. The training hyperparameters are presented in Table 2.

Table 2. Hyperparameters used in the CNN model

Training Parameters					
Regularization	**Batch Size**	**Optimizer**	**Loss function**	**No. of Epochs**	**Metrics**
L2 (0.01)	32	Adam	Binary Cross-entropy	20	Accuracy

With each training cycle, the model was able to continuously adjust its parameters to minimize errors and improve accuracy and learned to identify key stroke-related features in CT scans. The layer architecture of the proposed CNN is given in Table 3.

Table 3. Details of the proposed classifier architecture.

Layer	**Output Shape**	**Parameters**
Conv2D	(None, 126, 126, 8)	224
MaxPooling2D	(None, 63, 63, 8)	0
Conv2D	(None, 61, 61, 32)	2,336
BatchNormalization	(None, 61, 61, 32)	128
MaxPooling2D	(None, 30, 30, 32)	0
Conv2D	(None, 28, 28, 64)	18,496
MaxPooling2D	(None, 14, 14, 64)	0

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Layer	Output Shape	Parameters
Conv2D	(None, 12, 12, 64)	36,928
BatchNormalization	(None, 12, 12, 64)	256
MaxPooling2D	(None, 6, 6 64)	0
Flatten	(None, 2,304)	0
Dense	(None, 128)	295,040
Dense	(None, 64)	8,256
Dense	(None, 1)	65

The overall methodology along with the integrated pipeline has been depicted in the below figure, highlighting the major stages starting from data preprocessing, to model evaluation and final classification.

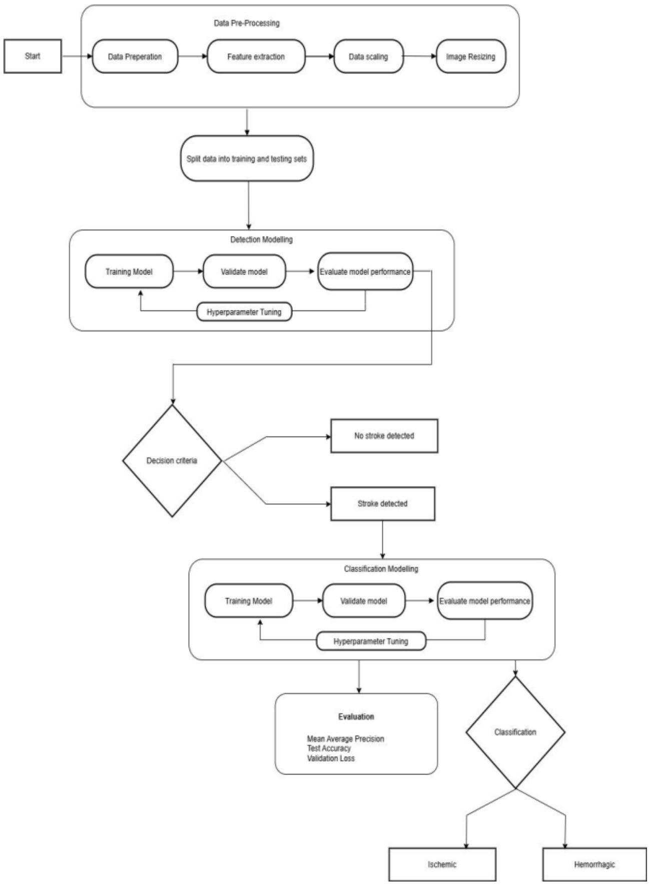

Figure 1. Pipeline for the Stroke Detection and Classification System

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Figure 1. illustrates the high level pipeline for the proposed system using deep learning approach. The workflow begins with data pre-processing of the patient health records and CT scans. The images are prepared, resized, and standardized to improve model, and performance ensure consistency. Once the data is ready, it's divided into training, testing and validation sets. The first stage of model training focuses on stroke detection, where the system determines whether a stroke is present based on demographic and health related parameters. If no stroke, the process ends there. However, if stroke is predicted, the pipeline proceeds where a classification model is used to differentiate between ischemic and hemorrhagic strokes. Throughout the training and classification stages, performance is continuously evaluated, and hyperparameters are fine-tuned to optimize accuracy.

RESULTS

This section outlines the findings of the two proposed models that were trained for stroke detection and classification. The performance of these models are also drawn in comparison to the competing models which unfortunately, failed to achieve the expected benchmarks. Key evaluation metrics like accuracy, precision-recall, AUC-ROC scores highlight their strengths and limitations of each approach [17].

Multilayered Perceptron Network

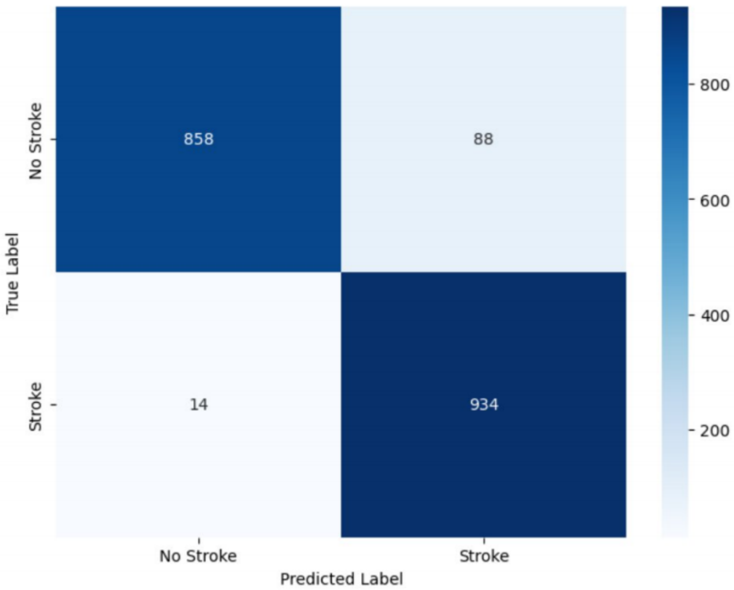

Figure 2. Confusion Matrix of the trained MLP

Figure 2. presents the confusion matrix for the trained MLP model, provides a clear understanding of its classification performance by illustrating the distribution of predicted and actual labels. The model correctly identified 858 non-stroke and 934 stroke cases, showing strong accuracy. However, it misclassified 88 non-stroke cases as stroke (false positives) and 14 stroke cases as non-stroke (false negatives). The high diagonal values indicate the model effectively distinguishes between stroke and non-stroke cases with minimal error.

Comparison with Random Forest

Prior to the MLP model, Random Forest model was trained on the preprocessed dataset. However, a minimal accuracy of 84% was obtained which was lower than expected. The below visualisations plot illustrate the performance differences of the Random Forest against the proposed MLP model.

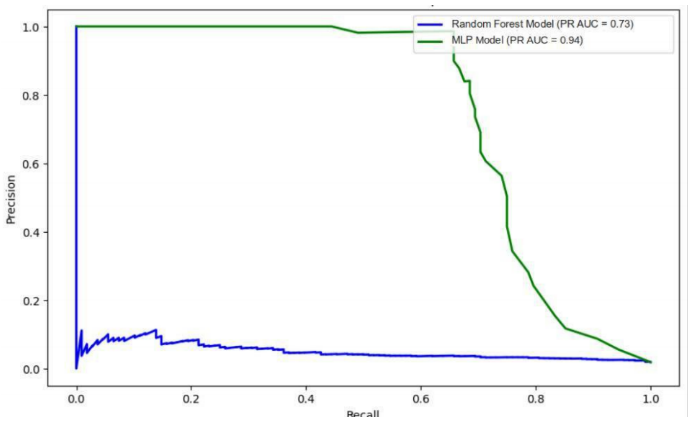

Figure 3. Precision Recall Curve Comparison

Figure 3. presents the precision recall analysis which contrasts the performance of the Multi-Layer Perceptron (MLP) against the Random Forest algorithm. The MLP algorithm achieves a higher area under the precision - recall curve, with value of 0.94, which indicates superior performance in predictability against that of Random Forest which only yields an AUC of 0.73. The MLP has high precision throughout the recall levels showing its resilience when processing datasets which have class imbalances. Inversely, the Random Forest model has a sharp drop indicating reduced capability in accurately identifying true positives cases. These findings show the ability of the MLP to produce better results.

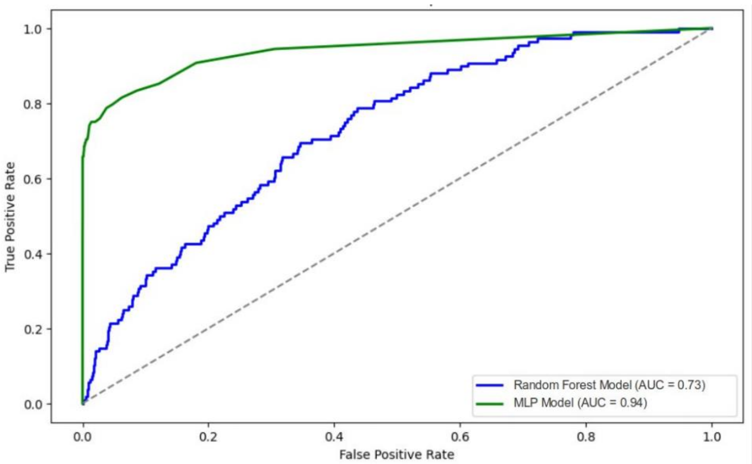

Figure 4. ROC Curve Comparison

Figure 4. presents the ROC curve comparing the MLP and Random Forest models. The MLP model has a higher AUC of 0.94, indicating strong classification performance, while the Random Forest model scores 0.73. The MLP curve stays close to the top-left corner, reflecting high true positive rates with only few false positives. In contrast, the Random Forest curve is less steep, indicating comparatively weaker discrimination. These results highlight the superior performance of the MLP model in making accurate predictions.

Convolutional Neural Network

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

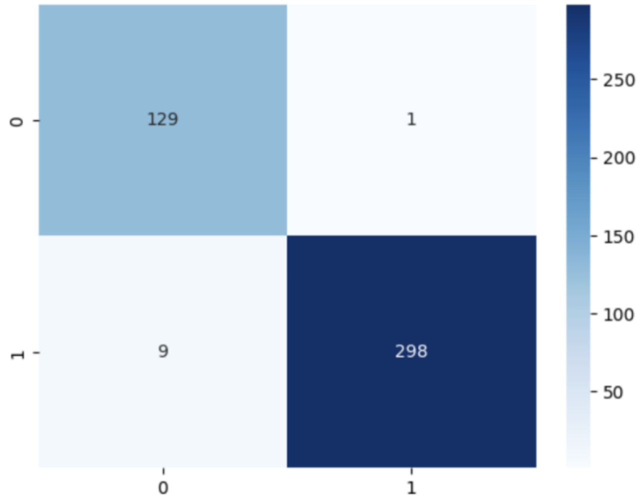

Figure 5. Confusion Matrix for the trained CNN

Figure 5. depicts the confusion matrix for the Convolutional Neural Network (CNN) model,which gives an overview of the classification outcomes of the model [18]. The matrix has four key metrics: the top left is the accurately identified true negative instances, the top right is the incorrectly identified positives, the bottom left is the inaccurately identified negatives and the bottom right is the accurately identified true positives. The model classified 129 instances with a negative which identified ischemic stroke correctly and 298 positive instances which indicate correct detection of hemorrhagic strokes. The model only produced 1 false positive and 9 false negatives. The high concentration of accurately identified true positives, and true negatives shows the model's accuracy in classifying the CT scans into ischemic and hemorrhagic with a high accuracy.

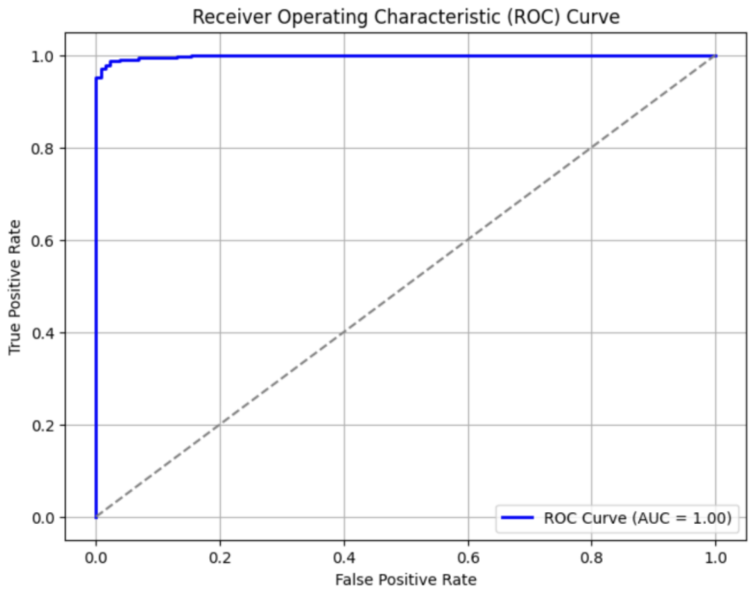

Figure 6. Receiver Operating Characteristic (ROC) curve [19] for trained CNN model

Figure 6. illustrates an ROC curve, visually representing the CNN's ability to differentiate between positive, and negative classes . The curve plots the true positive rate versus the false positive rate, depicting how well the model separates the two categories. A model with perfect discrimination achieves Area Under the Curve (AUC) of 1.00, as seen in this case, which indicates exceptional performance. The near-vertical rise and plateau along the upper boundary suggest that the model achieves high true positive rate, while maintaining near-zero false positive rate. The dotted diagonal represents a random classifier, and the fact that our model's curve stays far above it confirms that it significantly outperforms random guessing.

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

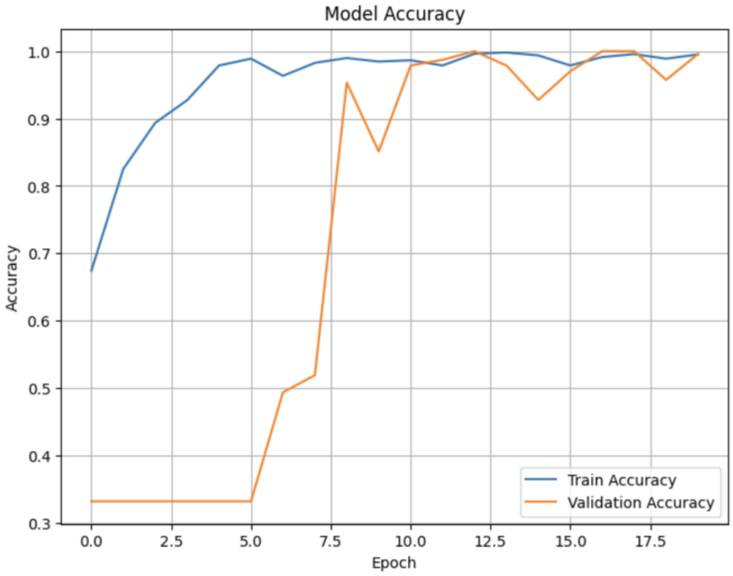

Figure 7. Training, and validation accuracy curves of the CNN model

Figure 7. shows an accuracy plot of the CNN model's learning behaviour over 20 training epochs. The blue curve represents the training accuracy, while the validation accuracy is represented by the orange curve. In early epochs, training accuracy rises quickly as the model learns patterns from the training data. However, the validation accuracy remains low at first, showing initial difficulty in generalizing to unseen data. Around the 6th epoch, the validation accuracy sharply increases that eventually aligns with training accuracy. This suggests that the model overcomes initial challenges, potentially due to improved feature extraction as training progresses. By the later epochs, both training and validation accuracies converge near 1.0, indicating that the model has achieved high predictive performance. The absence of a significant gap between the two curves in the final epochs suggests minimal overfitting. However, the sudden rise in validation accuracy around epoch 6 indicates an optimization breakthrough, due to parameter adjustments or a key feature becoming dominant in learning. Overall, the accuracy curve suggests that the CNN model has been trained effectively, achieving robust generalization capabilities.

Comparison with VGG and ResNet50

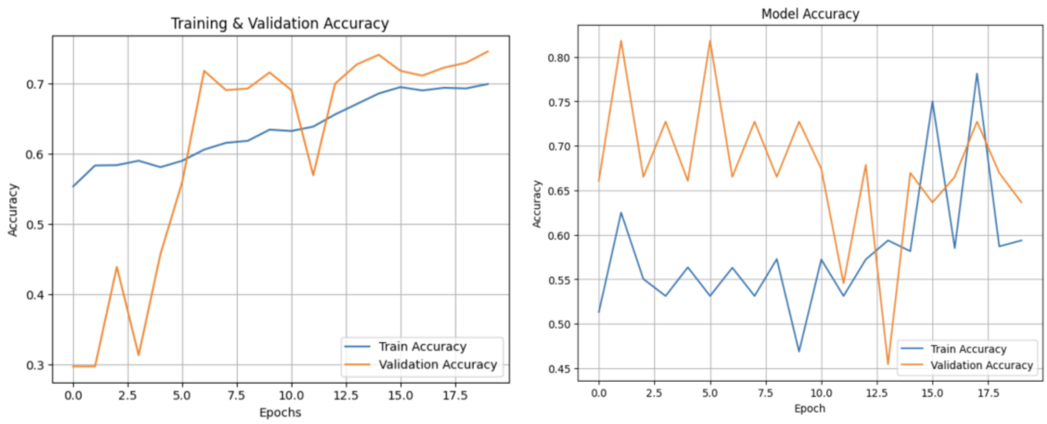

Figure 8. Training, and validation accuracy curves of the VGG model (left) and ResNet50 (right)

Figure 8. (left) illustrates the learning progression of the pre-trained VGG model over 20 epochs. The training accuracy represented by the blue line shows a smooth gradual increase indicating a stable learning process of the VGG without any sudden variations. The orange line shows the validation accuracy which has significant volatility during the initial few epochs indicating a gap in the model's ability to adapt to unseen data. This suggests that although VGG has pre-existing weights and offers a good starting point, it may still not be optimally aligned with the

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

specific dataset provided resulting in an initial volatility during learning. The training accuracy furthermore shows a relatively upward trajectory and remains under expected levels when compared to the variability observed in the validation accuracy. This delta could suggest that the model may not be obtaining the optimal feature representations for the provided dataset. The validation accuracy exceeds the training accuracy over several epochs, which points to the potential of over-fitting of the pre-trained VGG model.

Figure 8. (right) illustrates the training, and validation curves for the ResNet Model. The accuracy curve shows significant variability contrary to the expected smooth convergence that is characteristic of a deep neural network, particularly within the validation set. The validation accuracy exhibits a variable and unpredictable back and forth between 45% to 80%, suggesting the model is having difficulty in generalizing unseen data. The training accuracy also exhibits inconsistency and lacks a consistent upward trend, which suggests that the model may not be effective at learning resilient feature representations. These inconsistencies raise concerns regarding the model's adaptability.

Both models showcased in high performance, displaying their potential for clinical use. The MLP achieved 94.67% accuracy in stroke prediction using patient attribute data like age, BMI, and glucose levels. CNN achieved 98.6% accuracy in CT scan classification. These results thereby confirm the effectiveness and reliability of our methods in both stroke prediction and classification.

CONCLUSION

This paper details our research in which we introduced two distinct methods for brain stroke detection. Each of the methods employ a deep learning strategy which is tailored to that specific type of data. For the patient health metrics we utilized a Multi-Layer Perceptron (MLP) to evaluate stroke probability and for classification of stroke using CT Scans a Convolutional Neural Network (CNN) was used [18]. Key advantage of the MLP model was its ability to manage class imbalances which is common in medical metric data. By using SMOTE and careful selection of the features, the model balanced the given dataset which ensured good predictions accuracy across all types of data. This enhanced the model's F1 score, showing its ability to generalize across different patients' medical data and regularization methods, such as dropout, batch normalization and weight decay, which prevented overfitting of the model, and improved the model's ability to handle new data and unseen data. These methods rendered the MLP accurate and reliable for clinical data sets that could have variable levels of complexity. The success of the MLP with the patient attribute data indicates the utility of employing health attributes in prediction and CNN's capability in precise image classification proving the potential of deep learning in medical imaging. Together, the MLP can be employed as preliminary screening for stroke and the CNN for confirmation of diagnosis. The most important parts were data preprocessing and augmentation. SMOTE in MLP and data augmentation in CNN handled the data imbalance and variability. These techniques are enhanced in their accuracy and reliability in real world applications. Several regularizations such as the dropout method and batch normalization transformed the model into something clinically applicable. Finally, our study results show the efficacy of deep learning models in facilitating the process of stroke detection and classification. Both MLP and CNN models showed highly accurate and dependable output in stroke prediction from clinical medical data and brain CT-scan image classification respectively. The effectiveness of these models can be a blend of the sophisticated algorithms, best data management, and many regularization techniques rendering them very promising for real-world use. Future studies may explore incorporating these systems into a single system for stroke detection and classification and modify them to other medical conditions and any other variable which may further increase their usage.

DISCUSSION AND FUTURE PROSPECTS

The outcome of our work creates new avenues for further studies and application in the area of medical diagnostics. The MLP and CNN models had high potential in improving the stroke diagnosis, and this would lead to earlier diagnosis and more precise diagnosis and immediate medical treatment.It goes without saying that early intervention is key for stroke cases, as it improves outcomes when patients receive treatment in a timely manner.These models would be integrated into the clinic workflows where they would act as accurate decision-making systems for healthcare providers. For example, the MLP model would be used in primary healthcare clinics to screen high-risk patients according to their health profiles, and the CNN model would help radiologists interpret brain CT scans more accurately and with fewer inputs.

Aside from stroke detection, the developed methods are highly versatile and could be utilized for diagnosing other medical illnesses. For instance, the feature selection and data cleansing steps of the MLP model may be used to forecast heart disease or diabetes, just as the CNN architecture could be used for detecting brain tumors, hemorrhages, or other abnormalities in the medical imagery. This adaptability highlights the application of machine learning in healthcare and its potential solutions towards various medical challenges [19]. Additionally, the use of these models in AI-integrated healthcare systems could change the world for the better by analyzing medical data to offer healthcare predictions and diagnoses that are done automatically and which improves patient care.

Still, in spite of their promising outcomes, the models have certain restrictions. Their reliance on the training data's quality and heterogeneity is one of the main hurdles. It is possible that both the MLP and CNN models will have encountered so-called 'learning limitation' problems with data from other populations or medical systems, particularly if the training data is too narrow. To address this, future work could focus on deploying the models on a greater variety of datasets which include wider ranges of patients, regions, and clinical practices. For the CNN model, transfer learning could be valuable. In this strategy, previously trained models on similar datasets can be used for aid because they make the data more usable and cut back on the extensive training data that is needed.

Another limitation of the MLP model includes the lack of ability to generalize the results, which lead the model to overfitting the data provided when using a small range of features or reduced number of data samples. Although techniques such as batch normalization and dropout helped mitigate this issue, there is still room for improvement. The model's performance would likely improve with increased reliability on unseen data through techniques such as clinical data specific augmentation, cross validation, and early stopping. Furthermore, the model would benefit from a more representative training set by broadening the dataset with additional health parameters or by multi-institutional collaboration.

The CNN model, despite its high accuracy, has associated challenges largely related to its computational complexity. Training and fine-tuning require considerable computational power which restricts acceptance of the model in many regions, particularly those that are low in resource. This could be improved by performing further research into the use of MobileNet or EfficientNet architectures as these are known to have a good trade off between computation and performance. The model could also be made less complex by pruning redundant or less important neurons while maintaining accuracy.

Both models have a common, essential problem that needs to be solved: their lack of interpretability which makes it difficult for medical professionals to use the models. CNNs are often seen as "black boxes". Therefore, clinicians may distrust predictions without insight into the model's decisions. Techniques like SHAP(SHapley Additive exPlanations), or Grad-CAM(Gradient-weighted Class Activation Mapping) can explain model decisions, improving transparency and trust in predictions.For instance, Grad-CAM can identify regions of brain scans on which the CNN model focuses the most while classifying. This is similar to the way SHAP values can determine the contribution of each health parameter in the MLP model's results, thereby making its results interpretable and easier to understand.

In conclusion, even though MLP and CNN models presented in this study show promise for stroke detection and classification, addressing their limitations is critical for successful deployment in the real-world healthcare settings. These models will better serve the clinical requirements by improving generalization, regularization for robustness, using lightweight architectures for complexity reduction and enhancing interpretability.There needs to be further study of the application of these models in integrated diagnosis systems incorporating more than one source of data, like electronic health records and medical images, to provide an integrative model of patient treatment. In doing so, these models can introduce a revolutionary factor towards improving the health status, and the profession of medical diagnosis [13].

REFERENCES

[1] A. Srinivas, Joseph Prakash Mosiganti, A brain stroke detection model using soft voting based ensemble machine learning classifier, Measurement: Sensors, Volume 29, 2023, 100871, ISSN 2665-9174, https://doi.org/10.1016/j.measen.2023.100871.

[2] Muhammad Irfan, Abdulhamit Subasi, Noman Mustafa, Tomi Westerlund, Wei Chen, Chapter 5 - An evaluation of pretrained convolutional neural networks for stroke classification from brain CT images, Editor(s):

Abdulhamit Subasi, In Artificial Intelligence Applications in Healthcare and Medicine, Applications of Artificial Intelligence in Healthcare and Biomedicine, Academic Press, 2024, Pages 111-135, ISBN 9780443223082, https://doi.org/10.1016/B978-0-443-22308-2.00003-2.

[3] Igor Kononenko, Machine learning for medical diagnosis: history, state of the art and perspective, Artificial Intelligence in Medicine, Volume 23, Issue 1, 2001, Pages 89-109, ISSN 0933-3657, https://doi.org/10.1016/S0933-3657(01)00077-X.

[4] P. Gahiwad, N. Deshmane, S. Karnakar, S. Mali and R. Pise, "Brain Stroke Detection Using CNN Algorithm," 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), Lonavla, India, 2023, pp. 1-4, doi: 10.1109/I2CT57861.2023.10126125.

[5] M. A. Saleem et al., "Innovations in Stroke Identification: A Machine Learning-Based Diagnostic Model Using Neuroimages," in IEEE Access, vol. 12, pp. 35754-35764, 2024, doi: 10.1109/ACCESS.2024.3369673.

[6] S. Mushtaq and K. S. Saini, "A Review on Predicting Brain Stroke using Machine Learning," 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2023, pp. 667-673.

[7] B. R. Gaidhani, R. R.Rajamenakshi and S. Sonavane, "Brain Stroke Detection Using Convolutional Neural Network and Deep Learning Models," 2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT), Jaipur, India, 2019

[8] F. H. M. Mohammed, A. Ba Alawi, F. Bozkurt and M. D. Geyikoğlu, "An Efficient Detection Model for Brain Stroke Based on Transfer Learning," 2024 4th International Conference on Emerging Smart Technologies and Applications , Sana'a, Yemen, 2024, pp. 1-5, doi: 10.1109/eSmarTA62850.2024.10638910.

[9] Shinde, S., Kurhekar, M. P., Diwan, T., Pikle, N. K., & Gulhane, M. (2024). Design of a Novel Enhanced Machine Learning Model for Early Prediction of Cerebral Stroke. International Journal of Computing and Digital Systems, 15(1), 1807-1821. https://dx.doi.org/10.12785/ijcds/1501127.

[10] Rahman, S., Hasan, M. and Sarkar, A.K. 2023. Prediction of Brain Stroke using Machine Learning Algorithms and Deep Neural Network Techniques. European Journal of Electrical Engineering and Computer Science. 7, 1 (Jan. 2023), 23–30. DOI:https://doi.org/10.24018/ejece.2023.7.1.483.

[11] Gupta, Sarthak and Jadaun, Akhalesh Singh and Gupta, Pranshu and Larhgotra, Anamika, 'NeuroDetect: A Machine Learning Approach for Early Detection of Brain Stroke' (June 4, 2024). Available at SSRN: https://ssrn.com/abstract=4854167orhttp://dx.doi.org/10.2139/ssrn.4854167

[12] Alruily, Meshrif, Sameh Abd El-Ghany, Ayman Mohamed Mostafa, Mohamed Ezz, and A. A. Abd El-Aziz. 2023. "A-Tuning Ensemble Machine Learning Technique for Cerebral Stroke Prediction" Applied Sciences 13, no. 8: 5047. https://doi.org/10.3390/app13085047

[13] H L, G., Flammini, F., Kumar V, R., & N S, P. (Eds.). (2025). Recent Trends in Healthcare Innovation: Proceedings of the Annual International Conference on Recent Trends in Healthcare Innovation (AICRTHI 2024), Mysuru, India, October 24th–25th, 2024 (1st ed.). CRC Press. https://doi.org/10.1201/9781003501367

[14] Dubey Y, Tarte Y, Talatule N, Damahe K, Palsodkar P, Fulzele P. Explainable and Interpretable Model for the Early Detection of Brain Stroke Using Optimized Boosting Algorithms. Diagnostics. 2024; 14(22):2514. https://doi.org/10.3390/diagnostics14222514

[15] H.Georg Schulze, L.Shane Greek, Boris B. Gorzalka, Alan V. Bree, Michael W. Blades, Robin F.B. Turner, Artificial neural network and classical least-squares methods for neurotransmitter mixture analysis, Journal of Neuroscience Methods, Volume 56, Issue 2, 1995, Pages 155-167, ISSN 0165-0270, https://doi.org/10.1016/0165-0270(94)00118-Z.

[16] Raj, P., Sundaravadivazhagan, B., Raja, A.S., & Alani, M.M. (Eds.). (2025). Edge AI for Industry 5.0 and Healthcare 5.0 Applications (1st ed.). Auerbach Publications. https://doi.org/10.1201/9781003442066

[17] "Enhancing Cardiovascular Disease Risk Prediction Using Resampling and Machine Learning", *International Journal of Innovative Science and Technology (IJIST)*, ISSN: XXXX-XXXX, Vol. 6, Issue 2, page no. 514-531, May-2024.

[18] Gururaj H L, Francesco Flammini, S Srividhya, Chayadevi M L, & Sheba Selvam (Eds.), 1st ed., CRC Press, 2024,"Proceedings of the 1st International Conference on Computing and Intelligent Information Systems (ICCIIS 2024), Bangalore, India, 19-20th April, 2024, Volume 1". https://doi.org/10.1201/9781032711157

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

[19] Anurag Tiwari & Manuj Darbari (Eds.), 1st ed., CRC Press, 2025",Emerging Trends in Computer Science and Its Application", https://doi.org/10.1201/9781003606635.
[20] Gupta, M., Meghana, P., Reddy, K.H., Supraja, P. (2023). Predicting Brain Stroke Using IoT-Enabled Deep Learning and Machine Learning: Advancing Sustainable Healthcare. In: Whig, P., Silva, N., Elngar, A.A., Aneja, N., Sharma, P. (eds) ICSD 2023. https://doi.org/10.1007/978-3-031-47055-4_10
[21] N. Victor, S. Bhattacharya, P. K. R. Maddikunta, F. M. Alotaibi, T. R. Gadekallu and R. H. Jhaveri, "FL-PSO: A Federated Learning approach with Particle Swarm Optimization for Brain Stroke Prediction," 2023 IEEE/ACM 23rd International Symposium on Cluster, Cloud and Internet Computing Workshops (CCGridW), Bangalore, India, 2023, pp. 33-38, doi: 10.1109/CCGridW59191.2023.00020.
[22] M. Mujahid, N. Ayesha, A. T. Azar, T. Saba and Z. Haider, "Human Brain Stroke Prediction using Machine Learning Methods with Synthetic Minority Oversampling Approach," 2024 Seventh International Women in Data Science Conference at Prince Sultan University (WiDS PSU), Riyadh, Saudi Arabia, 2024, pp. 84-89, doi: 10.1109/WiDS-PSU61003.2024.00031.
[23] M. Anuprabha, K. Madhumitha, K. Sowmyadevi, T. Jaspar Vinitha Sundari and K. Paramasivam, "Prediction of Stroke Severity in Brain Through Machine Learning," 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2023, pp. 461-467, doi: 10.1109/ICICCS56967.2023.10142640.
[24] Nikita and G. Parashar, "Brain Stroke Detection and Prediction Using Machine Learning Approach: A Cloud Deployment Perspective," 2023 International Conference on Circuit Power and Computing Technologies (ICCPCT), Kollam, India, 2023, pp. 1705-1714, doi: 10.1109/ICCPCT58313.2023.10245699.
[25] Fernandes, J. N. D., Cardoso, V. E. M., Comesaña-Campos, A., & Pinheira, A. (2024). Comprehensive review: Machine and deep learning in brain stroke diagnosis. Sensors (Basel), 24(13), 4355. https://doi.org/10.3390/s24134355
[26] "Brain stroke detection using convolutional neural networks", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.8, Issue 10, page no.e304-e307,