

DEEP LEARNING-BASED BRAIN STROKE DETECTION: CLINICAL RELIABILITY ANALYSIS OF CNN AND HYBRID NEURORESUNET ARCHITECTURES

Anil Hingmire¹, Sunayana Jadhav², Megha Trivedi³, Trupti Shah⁴, Hemant Appa Tirmare⁵

¹ Department of Computer Engineering, Vidyavarhini's College of Engineering and Technology, Vasai, Mumbai University, Maharashtra, India,

Email: anil.hingmire@vcet.edu.in

² Department of Electronics and Telecommunication, Vidyavardhini's College of Engineering and Technology, Vasai, Mumbai University, Maharashtra, India, sunayana.jadhav@vcet.edu.in

³ Department of Computer Engineering, Vidyavardhini's College of Engineering and Technology, Vasai, Mumbai University, Maharashtra, India, megha.trivedi@vcet.edu.in

⁴ Department of Electronics and Telecommunication, Vidyavardhini's College of Engineering and Technology, Vasai, Mumbai University, Maharashtra, India, trupti.shah@vcet.edu.in

⁵ Department of Computer Science and Technology, School of Engineering and Technology, Kolhapur, Shivaji University, Maharashtra, India, hat_tech@unishivaji.ac.in

Corresponding Author: Anil Hingmire (Email: anil.hingmire@vcet.edu.in)

Abstract: Brain stroke is a leading cause of death and long-term disability in the world and timely and accurate diagnosis is critical to better patient outcomes and decreasing mortality rates. The automated detection of stroke using deep learning has been a new promising way to assist clinical decision making and improve the efficiency of diagnosis. Two deep-learning models: a Convolutional Neural Network (CNN) and a Hybrid NeuroResUNet model, are compared in this study for brain stroke automated prediction by computed tomography (CT) scan images. A labeled dataset consisting of stroke and non-stroke CT images was used for training and evaluation of the models. Standard Evaluation metrics such as accuracy, precision, recall, specificity, F1 score and confusion matrix analysis were used to evaluate performance. The experimental results showed that the Hybrid NeuroResUNet outperforming the other methods with an overall accuracy of 95.06%; however, the method had a relatively high false-negative rate which could pose risks for clinicians using the system for stroke detection. The CNN model showed better accuracy of 93.94%, sensitivity of 96.77%, specificity of 91.43%, and F1 score of 93.75%, compared to the others. The CNN model's performance in accurate classification of stroke and non-stroke cases was also verified using a confusion matrix analysis. The results showed that although the CNN model had a slightly lower accuracy, it was more clinically reliable and robust in automatic brain stroke detection and thus proved to be more suitable for clinical applications..

Keywords: Brain Stroke Prediction, Convolutional Neural Network, Hybrid NeuroResUNet, Stroke Classification, Deep Learning, Medical Image Classification.

1. INTRODUCTION

A stroke occurs when blood ceases to flow through one of the arteries, or one of the blood vessels breaks open, causing the brain tissue to not receive enough blood, oxygen, and nutrients. This disturbance can quickly lead to irreversible brain damage and therefore, early treatment is essential to get a good outcome. There are different types



of strokes and recovery from stroke depends on the type. However, prompt identification of stroke symptoms and effective treatment of stroke risk factors reduces stroke mortality and long-term disability. Ischemic stroke accounts for ~87% and intracerebral and subarachnoid hemorrhage account for ~10% and 3%, respectively, according to the WHO [1]. Strokes are typically divided into ischemic or hemorrhagic; almost 90% are ischemic, a result of abnormally low blood flow to the brain [2].

The pathogenesis of stroke involves oxidative stress and neuroinflammation. Cellular injury results from the imbalance of reactive species and antioxidant defenses, and further cellular injury occurs from neuroinflammation, which is an immune activation and release of cytokines. These pathophysiological mechanisms collectively play an important role in the severity of tissue damage, the subsequent repair mechanisms, and clinical outcomes after stroke, and are therefore very relevant targets for therapy [3]. CTP, DWI and PWI advanced imaging can be used to select AIS patients for EVT up to 6-24 hrs after onset by measuring the infarct core and penumbra and defining the target mismatch criteria. This reduces reliability due to artefacts, differences among vendors and estimation errors in core volume. Some recommendations rely on the non-contrast CT ASPECTS and CTA; other recommendations are pro advanced imaging to obtain tighter guidelines. The use of conventional and advanced imaging has been recently shown to optimize patient selection for EVT [4].

Advanced neuroimaging (CT) is important for markers for disease management. These methods allow an important assessment of the site and extent of the lesion, especially in the acute stage of stroke, and the assessment of brain structure and function can be performed throughout the entire cortex, not only in the vicinity of the lesion [5]. Due to diverging and highly individual processes of recovery, it is hard to predict gait recovery following stroke. Traditional clinical, neurophysiology and neuroimaging are informative but not sufficiently accurate for prediction of outcomes.

This study makes the following main contributions:

1. An automated brain stroke detection system is proposed based on Hybrid NeuroResUNet, which combines the deep feature learning ability of ResNet and the localization power of U-Net for a better analysis of neuro images.
2. To make a comprehensive comparative study of Hybrid NeuroResUNet and conventional CNN architectures, the publicly available brain stroke imaging dataset is used and clinically relevant evaluation metrics such as Accuracy, Precision, Recall and F1-score are used.
3. This study shows that the overall classification accuracy is not sufficient and the importance of stroke specific recall, sensitivity and reduction of false negative for clinically valid stroke diagnosis.
4. To assess the robustness, generalisation and clinical applicability of the developed models, an extensive performance assessment including confusion-matrix analysis, cross-validation, statistical significance testing and error analysis is performed.
5. Through experimental results, it is found that balanced detection performance is more important than marginal gains in the overall accuracy, which will be helpful for the design of reliable AI-supported stroke diagnosis systems.

The paper is organized into five sections. Section 1 introduces the research problem and objectives. Section 2 reviews the relevant literature on brain stroke detection and classification. Section 3 presents the proposed methodology, including data preprocessing, model architecture, and experimental setup. Section 4 discusses the experimental results and comparative performance analysis. Finally, Section 5 concludes the paper and outlines potential directions for future research.

2. RELATED WORK

In this research, Genevieve Richarda et al investigated if brain age measured by MRI could predict cognitive performance and computerized cognitive training response following stroke. Brain age may well be reliably estimated in stroke patients, but there was no association with either cognitive outcome or responsiveness to training. Filling this gap, Emam Alanazi et al. (2021) have demonstrated that machine learning can predict stroke outcomes, with laboratory results, which are usually the basis of most studies. In their paper, they created a random forest of resampling of their data and obtained an accuracy of 96%, sensitivity of 97% and specificity of 96%. Their model was very well predictive in the lab setting and more studies on stroke subtypes are recommended to improve the prediction [7].

In 2022, a study done by Nitish Biswas et al. to predict stroke data in imbalanced data sets with the help of Random Over Sampling and improve the performance of the classifiers. The data balancing and careful tuning of the classifiers is very important as was established when the accuracy of four classifiers was measured and it was found that more than 96% of the data were accurate with Support Vector machines and Random Forest having nearly perfect accuracy [8]. Vamsi Bandi et al. proposed the SPN algorithm, an advanced random forest-based framework that could assess the risk due to different types of strokes. The accuracy of the corresponding SPR model was found to be an amazing 96.97% and better than the state-of-the-art methods [9].

Md. Monirul Islam et al. suggested a stroke prediction system based on a Random Forest, trained with a 5,110 sample and 12 features. This model achieved 96% on the total accuracy, precision, recall, and the overall performance measure using EDA as a preprocessing and feature selection, outperforming the Decision Trees model [10]. An ensemble stacking-based neural ischemia prediction model was proposed by Samaa A. Mostafa et al., which used SMOTE and data standardization for balancing the class. The stacking approach performed better than that of an individual classifier, with 97% accuracy and a 94% Matthews correlation coefficient (MCC) [11].

Rukiye Karakis et al. took the approach of combining demographic data with DTI to identify white matter damage and successfully predicted motor impairments with a three-dimensional residual convolutional neural network with accuracy of 92% [12]. A. Srinivas et al. worked on ensemble learning model using weighted classification probability which achieved 96.88% accuracy for predicting stroke events in UCI data [13]. Saleem et al. employed a genetic algorithm-Bi-LSTM algorithm for CT image and reported an accuracy of 96.5% [14]. In addition, MRI morphometry was used by Richarda et al. to predict brain age, where only associations with cognitive outcome were found as weak [6]. All these studies reflect progress in stroke prediction and point out the need for more robust models using multimodalities with accurate prediction of subtypes of stroke, as mentioned by Alanazi et al. [7].

Balanced data enhances the performance of a stroke-prediction model, so Nitish Biswas et al. suggest that more work needs to be done in the area of balancing and interpretability strategies [8]. According to Vamsi Bandi et al., the SPN predicts several types of stroke risks, but the assessment of risk for various types of strokes needs further improvement [9]. There are opportunities for feature augmentation/improvement of techniques to address the class imbalance issue as Md. Monirul Islam et al. presented that ensemble and advanced classifiers give high accuracy in predictions [10][11]. Detailed prediction can be supported by the investigation of other neuroimaging approaches. In the case of motor deficits, the performance of multimodal neuroimaging, like DTI in a multi-channel CNN, has been high [12]. Likewise, Muhammad Asim Saleem et al. proposed an early detection using a BiLSTM based on CT scan images, and A. Srinivas et al. proposed some ensemble methods which performed well but further improvement is required with more data sources and optimisation of models [13][14].

In recent years there have been some successes in hybrid deep learning for medical imaging. The CNN-LSTM model was also used by Hingmire et al., 2024, and achieved an accuracy of 94% of which there is potential for neuroimaging applications [16]. Hybrid UNet-ResNet architectures have been shown to improve sensitivity in detecting subtle ischemic lesions thanks to the attention mechanism [17]. The sensitivity of small lesions in MRI has been additionally improved by attention-guided frameworks, e.g., Triply Fusion Attention [18]. Some of the major restrictions are lack of annotation data and variation in lesions. The limitations of annotated data and lesion variations remain the main problematic issues in this work. Self-supervised learning can be seamlessly combined with multi-modal MRI fusion to achieve a comprehensive prognosis of stroke in the future. These limitations in incorporating multi-modal data, improving model generalization, and better prediction methods could be overcome through further research, leading to advances in stroke diagnosis and treatment.

In their paper, Sailasya et al. (2021) discussed the machine learning based methods for prediction of stroke using clinical and physiological data. Sailasya et al. found that the traditional classifiers such as Logistic Regression, Decision Tree, and Random Forest have moderate accuracy; Naive Bayes had a relatively good performance in structured data, but their method was based on non-imaging data only and they did not model space features. In other papers, improvements using deep neural networks and ensemble methods have been reported, but they tend to be more complicated and/or rely on handcrafted features. These challenges manifest themselves as a need for image based deep learning networks that can be trained to learn discriminative spatial patterns directly, using brain images, which is why convolutional models are used to provide more accurate and robust stroke detection [19].

Aniwat et al. (2022) developed a deep learning system that combines a multi-class segmentation system (DeepMedic) and a fine-tuned EfficientNet-B2 classifier for the identification of the subtypes and for measuring the lesion thickness and volume from head computed tomography (CT) images. The method is arranged structurally, but limited with specific datasets, hemorrhage sub types and missing other clinical signs (skull fracture revealed, midline

shift examined) to support the method, and further validation and clinical incorporation is needed [20]. Bonna Akter et al. (2022) suggested a machine learning framework to predict early brain stroke risk also using a structured clinical dataset of 5,110 patients with 12 attributes based on the structured dataset. They have used the methodology of data preprocessing, transformation, and standardization, and then training a variety of the classifiers, such as Random Forest, Support Vector Machine, and Decision Tree. Different measures of evaluation such as accuracy, sensitivity, error rate and log loss were used to assess the performance of models. The study showed that traditional machine learning models are effective in stroke prediction although it is limited by its use of tabular data as opposed to medical imaging and not on heterogeneous or incomplete datasets meaning that further work can be performed to enhance the robustness and generalization [21].

Azhar Tursynova et al. (2023) designed a computer-aided diagnostic system using a CNN to classify a brain stroke in the CT images into normal, ischemic and hemorrhagic strokes. The model achieved higher accuracy and higher recall compared to the traditional machine learning techniques, and it obtained 79 percent accuracy with data augmentation and early stopping technique. They conclude that deep learning techniques could be helpful for the automated detection of stroke and clinical decision making [22]. Xun et al. (2024) introduced an advanced U-Net based segmentation framework for analyzing brain MRI images, called ARGU-Net, which incorporated Residual Grouped Convolution Modules (RGCM), Convolutional Block Attention Modules (CBAM) and bilinear-Interpolation (BI) upsampling to enhance segmentation performance. The role of normalization techniques, loss functions, and network depth in the segmentation task were also explored in the model. Experimental results with the TCIA brain tumor MRI dataset were obtained and compared using the Dice coefficient, which achieved a score of 97.58% in comparison with the conventional models U-Net, ResUNet, SegNet and Mask R-CNN. It revealed that the residual grouped convolutions and attention mechanisms can improve feature extraction and boundary delineation in the medical image segmentation task. The framework was however tested only on the brain tumor MRI data from a single public dataset, and it was mainly used in a 2D segmentation scenario, which might not generalize well to other neurological diseases and 3D clinical imaging scenarios. Therefore, it must be tested on other datasets and volumetric medical images to ensure its robustness and applicability in the clinical setting [23].

Using clinical data, Gupta et al. (2025) compared several machine learning algorithms, such as logistic regression, support, and K-nearest neighbors, random forests, and neural networks to predict stroke risk. They discovered that the accuracy of their results was above 95% and the best performance was achieved by the neural networks. However, clinical data was reported as being poorly represented, data were small, and the classes were imbalanced, all of which were cited as limitations of the study for which the model could impact the generalizability and sensitivity to stroke, and improve the quality of clinical data and approaches that allow for better data imbalance [24]. Sailaja and Velmurugan (2025) suggested an automated brain stroke detection system using MRI images based on Deep Spiking MobileNet (DSpM-Net) framework. The method involved image denoising, lesion segmentation using R2U-Net, extraction of DT-LDNP features and Deep Spiking MobileNet to classify. The advanced feature extraction technique and the deep learning model achieved excellent results in the diagnosis of stroke from MRI images, with the TPR being 94.76%, the TNR being 91.87%, the accuracy being 92.85%, and the F1-scoring being 93.80%, demonstrating the potential of the proposed approach. The architecture does, however, depend on several pre-processing and manually designed steps for feature extraction, which can be computationally expensive, and limit the end-to-end learnability [25].

Inamdar et al. (2025) have presented attention-based deep learning model based on DAT, CAT and machine learning classifiers to classify brain stroke in computed tomography (CT) images with 99.51% accuracy. The model was optimized with the data from a single center, however, which may lead to over fitting and poor generalizability to other clinical settings [26]. Mostafapour et al. (2026) suggested a hybrid CNN-VGG16-XGBoost model that could detect brain stroke from MRI images with classification accuracy of 100% in less diagnostic time. The model has shown great promise but has only been tested in a small MRI database and has not been clinically validated in an external testing set, which has lowered the level of trust in the model's real-world use [27].

In this paper, Shakunthala et al. (2026) proposed BSD-IHSC-SMCNN framework to classify the three types of strokes (normal, ischemic and hemorrhagic) from the MRI images by means of image enhancement, segmentation, feature extraction and optimization. The model obtained the accuracy of 98%, 93%, and 96% for normal, ischemic and hemorrhagic stroke, respectively. It is, however, multi-stage (hence more complicated to compute), and it has only been tested with one MRI dataset, limiting it to generalizability to other clinical settings [28].

3. Materials and Methodologies

Although deep learning (e.g., CNNs and transfer learning) has improved the understanding of neuroimages, stroke detection is a relatively under-explored field, and subtle patterns are not detected by traditional methods. The advanced preprocessing was emphasized, along with hybrid deep learning models and robust evaluation. Image augmentation can be used to improve the ability of CNNs to capture stroke features, thereby improving diagnostic accuracy and speeding up the assessment.

The idea of an automatic stroke detection system using CT images is illustrated in figure 1. The Raw brain CTs are fed into the Input Block, followed by Image Preprocessing for noise reduction, normalization and contrast enhancement. The processed images are then fed to the Feature Extraction Block where the edges are detected, texture analysis is performed and numerical representation of the features is done. These features are fed into a Convolutional Neural Network for stroke pattern detection. The Automated Stroke Prediction Block classifies images as stroke-affected or normal and calculates the stroke probability. The results are evaluated in the Prediction Evaluation Block with performance metrics like accuracy and sensitivity; optimization feedback can be used to further tune the performance. Lastly, the status-normal or stroke-affected is predicted and delivered in the Output Block.

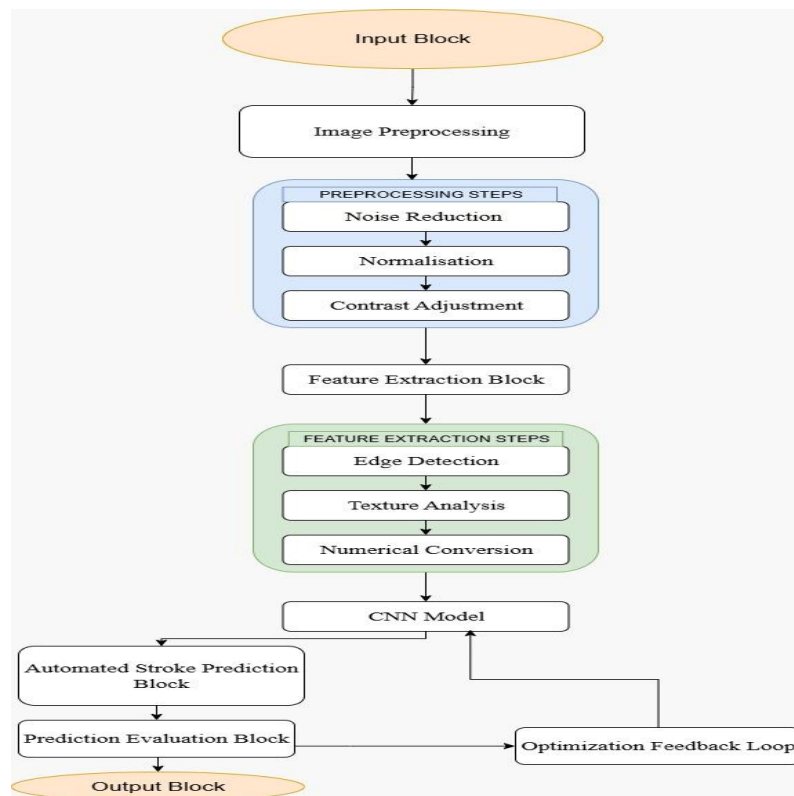


Figure 1. Stroke Detection System

The brain stroke prediction system workflow is shown in Figure 2, consisting of data collection, data preprocessing, training of the model, validation and prediction. It starts with collecting raw data from the sources relevant to the problem, then moves on to data preparation tasks like dealing with missing values, fixing inconsistencies, or finding patterns in the data that is available. This stage is used to make sure that the dataset is clean, complete, and ready for further analysis. The prepared data is then sampled and split into training and testing sets for the development and evaluation of the model.

To enhance the quality of training data and computational efficiency, additional data preprocessing steps are taken, including feature selection, feature scaling, data normalization, and dimensionality reduction. The techniques are used to extract the most relevant features while discarding noise and redundant features. The model is then trained with the selected features, and hyperparameter tuning is used to maximize the learning performance and predictive accuracy. Once they have been trained, post-processing is performed to improve the accuracy of the output produced and to increase the reliability in the predictions.

The trained model is then cross validated to ensure that it can be generalized and to avoid overfitting. The assessment of the performance uses accuracy, precision, recall, specificity, and F1-score, giving a comprehensive evaluation of the effectiveness of the predictions. Finally, the validated model is tested on test data sets or patients' input of new data sets and prediction accuracy, and reliability are monitored continuously for stroke prediction. Overall, it is a pipeline for the prediction of brain stroke, which is streamlined, interpretable, practically deployable and useful for brain stroke clinical decision support.

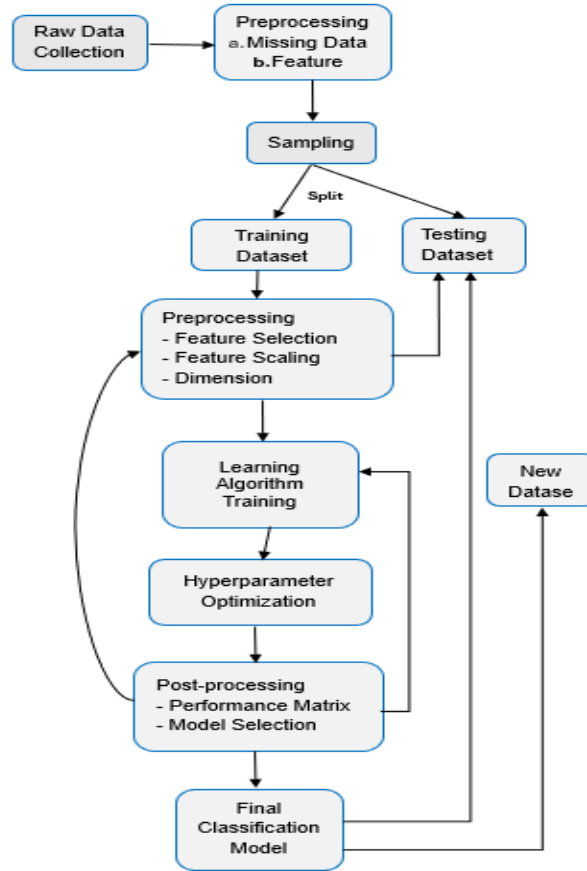


Figure 2. System Flowchart

3.1 Data Collection

Data collection forms the backbone of any serious ML model aiming to interpret medical images. In this work, 5,185 CT scan images from Kaggle [15] are used, which were chosen with quality, availability, and relevance in mind. The dataset was divided into training and testing/validation to ensure that evaluation will be done on unseen data: 4,148 images (80%) for training and 1,037 images (20%) for testing/validation.

3.2 Hybrid NeuroResUNet Model

Figure 3 shows the hybrid NeuroResUNet architecture that combines ResNet and U-Net approaches for improving brain stroke prediction by a combination of sequential classification and segmentation. The workflow starts with a brain CT that is pre-processed using a wavelet transformation to extract relevant spatial and frequency-based information to represent it in a better manner.

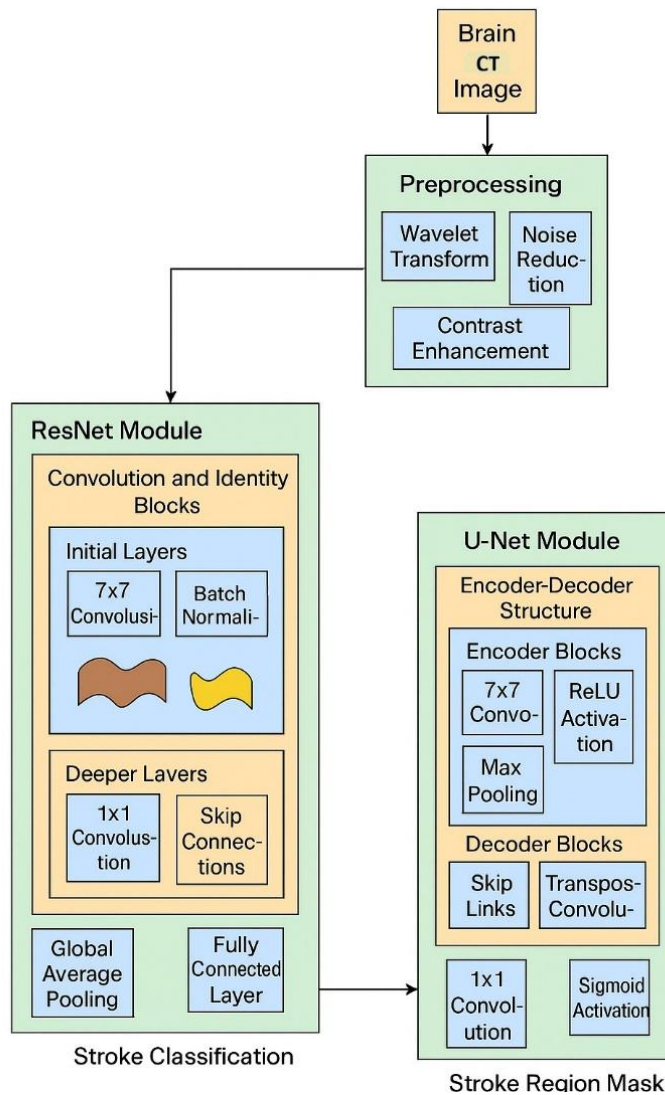


Figure 3. Hybrid NeuroResUNet Model

The preprocessed image is fed into a ResNet-based CNN for stroke detection. ResNet relies on residual connections to ensure gradient flow, thus allowing it to train deep models. It starts by having a 7×7 conv, batch norm, ReLU, and max pooling; then adding 1×1 and 3×3 residual blocks to capture hierarchical features; and finally, global average pooling and softmax FC to output the stroke class.

If stroke is detected, a U-Net segmentor refines the result. U-Net relies on an encoder–decoder with skip connections: the encoder first applies 3×3 convs and max pooling to extract features while the decoder applies transposed convolutions to recover spatial resolution. A final 1×1 conv with sigmoid activation yields the stroke mask. NeuroResUNet combines ResNet for robust classification with U-Net for precise localization and provides a comprehensive framework for automated stroke detection and segmentation to support clinical decisions.

Convolutional Neural Network (CNN) Model

Figure 4 illustrates a Convolutional Neural Network (CNN) model for brain stroke prediction using CT scans. The pipeline begins with input brain images that undergo preprocessing steps such as resizing, normalization, and augmentation to improve training efficiency.

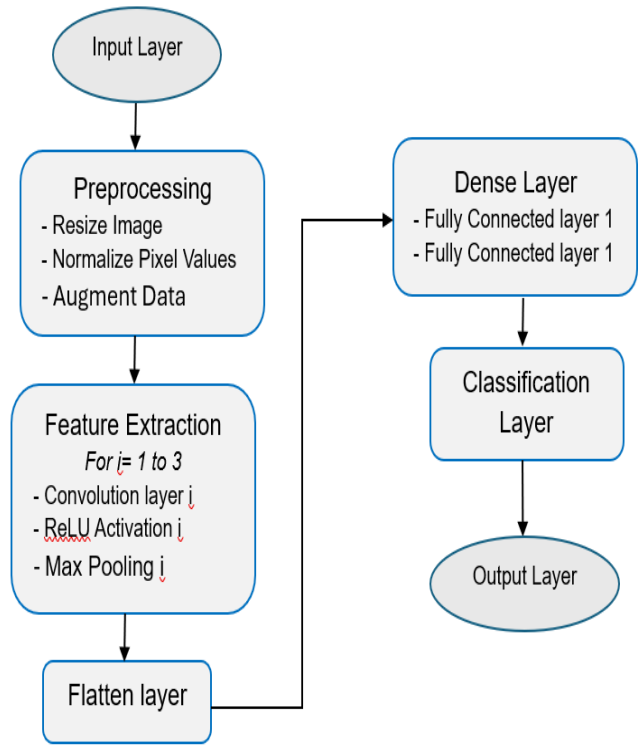


Figure 4. Block diagram of CNN Model

During feature extraction, edges are detected by convolutional layers using ReLU activation, while textures and patterns are reduced in their spatial dimensions by pooling layers without losing significant information. The resulting features are fed into fully connected layers, which model higher-level representations, followed by a classification layer using Softmax for determining the presence of stroke. The final output gives a prediction with an associated confidence score; thus, medical image analysis follows this sequence in stroke detection.

4. RESULTS AND DISCUSSION

This section compares the performance of a classic CNN with that of the proposed Hybrid NeuroResUNet model in predicting brain strokes using confusion-matrix-based metrics, namely accuracy, precision, recall, and F1-score. For both, Python implementations were made by utilizing TensorFlow and PyTorch, along with helper libraries such as NumPy, Pandas, scikit-learn, OpenCV, and Matplotlib for data processing, evaluation, and visualization. The CNN was realized by making use of the Keras API, while NeuroResUNet had a custom designed hybrid architecture. The Hybrid NeuroResUNet model was evaluated as a binary classifier-Class 0 (No Stroke) and Class 1 (Stroke)-whose precision, recall, F1-score, and support are described in Table 1.

Table 1. Classification Results of Hybrid NeuroResUNet

Class Name	Precision	Recall	F1-score
Class 0 (No Stroke)	1.0000	0.9506	0.9747
Class 1 (Stroke)	0.0471	1.0000	0.0899

The classification accuracy results of the Hybrid NeuroResUNet model for brain stroke prediction are given in Table 1. The model performed well in the No Stroke class, with a precision of 100%, recall of 95.06%, and an F1-score of 97.47%, which shows a good performance in identifying normal cases. These results, however, with a precision of 4.71% and a recall of 100% gave a low F1-score of 8.99% for the Stroke class. The findings show that the model was able to identify stroke patients but failed to provide a healthy number of false positive predictions, which led to poor performance of the model in making reliable stroke diagnosis.

Training Set			
TARGET \ OUTPUT	Class0 (No Stroke)	Class1 (Stroke)	SUM
Class0 (No Stroke)	1559 94.83%	0 0.00%	1559 100.00% 0.00%
Class1 (Stroke)	81 4.93%	4 0.24%	85 4.71% 95.29%
SUM	1640 95.06% 4.94%	4 100.00% 0.00%	1563 / 1644 95.07% 4.93%

Figure 5. Confusion Matrix Hybrid NeuroResUNet Model

The confusion matrix for brain stroke prediction model (Hybrid NeuroResUNet) is shown in figure 5. Confusion matrix of brain stroke prediction model (Hybrid NeuroResUNet) is presented in figure 5. The model achieved a recall of 95.06% for the class No Stroke with a correct classification of 1559 records from the total 1640 non-stroke. Furthermore, the number of the four predicted stroke cases was accurately categorised with a Precision of 100% for the Stroke class. However, the model misclassified 81 stroke cases as non-stroke, which resulted in a stroke recall of only 4.71%. As a result, the overall accuracy of the model was 95.07% but its performance on detecting stroke cases was flawed because of the high percentage of false negative cases (95.29%). The results presented here suggest that the model Hybrid NeuroResUNet has a good diagnostic performance for non-stroke cases but a low sensitivity in stroke recognition, limiting the clinical usefulness of the model for the reliable diagnosis of stroke.

```

127/127 ————— 29s 230ms/step - accuracy: 0.9918 - loss: 0.0269
127/127 ————— 4s 28ms/step - accuracy: 0.9649 - loss: 0.1331
127/127 ————— 74s 565ms/step - accuracy: 0.9743 - loss: 0.0771
Train Loss: 0.028011944144964218
Train Accuracy: 0.9897297024726868
-----
Validation Loss: 0.14130960404872894
Validation Accuracy: 0.9625360369682312
-----
Test Loss: 0.09855052828788757
Test Accuracy: 0.9763779640197754

```

Figure 6. Result of CNN Model

As it can be seen in Figure 6, the proposed CNN model of brain stroke prediction has very impressive results in terms of training, validation, and test databases. The accuracy of training at 98.97 percent and the low loss of training at 0.2801 indicates that the CNN model has learned the underlying pattern of the data very well. The validation accuracy of 96.25% and the low value of validation loss of 0.1413 show that the CNN model has minimum overfitting and, therefore, the proposed CNN model is generalized on unseen data. Moreover, with the test accuracy great at 97.64 and low-test loss at 0.0986, a good performance on entirely new data has separated stroke and non-stroke cases with a sharp line. This is also confirmed by the gradually dropping loss value in the training, validation and test datasets. Although CNN model has gained great precision, its generalization ability may be enhanced through various regularization methods, data augmentation and hyperparameter optimization. Altogether, the suggested model can be characterized as an extremely efficient and effective instrument in the process of categorizing strokes of the brain.

Figures 7 (a) and (b) represent the accuracy and loss of the CNN model of the training, validation, and testing data in a bar chart. Accuracy graph indicates high and reliable accuracy with training being 98.97, validation at 96.25 and test at 97.64 which means that the learning and generalization were excellent. Loss Graph represents low values of losses, which decline between 0.0280 (train) and 0.0986 (test), which justifies the strength and dependability of the model.

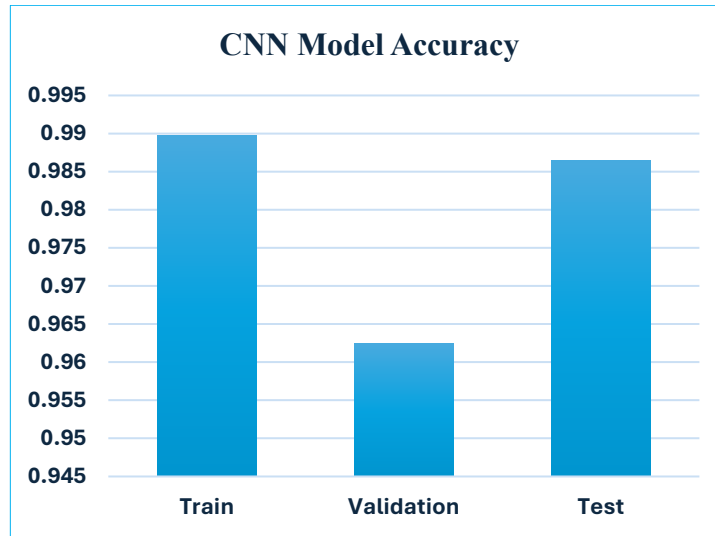


Figure 7 (a). CNN Accuracy (Training/Testing)

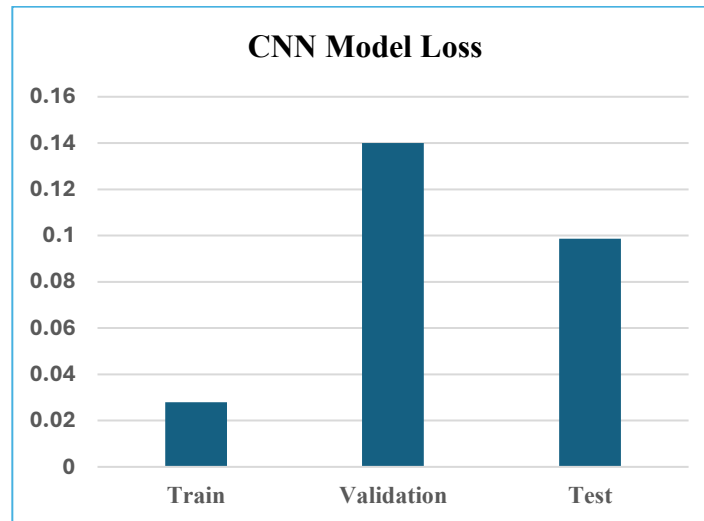


Figure 7 (b). CNN Loss (Training/Testing)

Table 2 shows the classification performance of the CNN model for brain stroke prediction. The model exhibited good discrimination power in both the classes. The model successfully identified the normal cases with a low false positive rate of 3.23% in the No Stroke class (Class 0) with precision 96.97%, recall 91.43% and F1-score 94.12%. Model performance for Stroke class (Class 1) achieved a precision of 90.91%, recall of 96.77% and F1 of 93.75%, with a relatively low false negative rate of 3.23% and very high sensitivity in detecting stroke cases. The specificity is also good: 96.77% in the No Stroke class and 91.43% in the Stroke class, further indicating the capacity of the model to differentiate between the two classes. The results demonstrate that the CNN model is able to provide balanced and reliable classification performance, which is suitable for supporting clinical stroke diagnosis where minimising the number of missed stroke cases is of vital importance.

Table 2. Classification Results of CNN Model

Class Name	Precision	Recall	FNR	F1-Score	Specificity (TNR)	FPR
Class 0 (No Stroke)	0.9697	0.9143	0.0857	0.9412	0.9677	0.0323
Class 1 (Stroke)	0.9091	0.9677	0.0323	0.9375	0.9143	0.0857

Table 3. Overall CNN Performance

Metric	Value
Accuracy	93.94%
Sensitivity (Stroke Recall)	96.77%
Specificity	91.43%
Precision	90.91%
F1-score	93.75%
Misclassified Samples	8 / 132

The overall performance of CNN model on the test data set is summarized in Table 3. Overall, the model was found to have an accuracy of 93.94%, with 124 out of 132 samples classified correctly. The sensitivity (recall) for stroke detection was 96.77%, suggesting the model was able to capture most of the instances of stroke and with low false negative rate. Its specificity of 91.43% shows this is a good recognition of non-stroke cases and has a low number of false-positive findings. Moreover, the model obtained a precision of 90.91% and the F1 score of 93.75%, indicating that the proposed model had a decent accuracy and precision. The CNN architecture was found to be robust and able to generalize well due to only eight samples being misclassified. To sum up, the CNN model is shown to be reliable and clinically useful for brain stroke prediction as it has achieved high sensitivity to detect stroke cases.

Training Set			
TARGET \ OUTPUT	Class0 (No Stroke)	Class1 (Stroke)	SUM
Class0 (No Stroke)	64 48.48%	2 1.52%	66 96.97% 3.03%
Class1 (Stroke)	6 4.55%	60 45.45%	66 90.91% 9.09%
SUM	70 91.43% 8.57%	62 96.77% 3.23%	124 / 132 93.94% 6.06%

Figure 8. Confusion Matrix of CNN model result

The confusion matrix was derived from CNN model using a training dataset for brain stroke classification and is shown in Figure 8. The matrix shows how many samples of each class (No Stroke, Class 0 or Stroke, Class 1) were classified correctly. A total of 124 samples were correctly classified out of 132 samples (64 No Stroke and 60 Stroke)

in the CNN model. Six non-stroke cases were incorrectly classified as stroke, while two stroke cases were misclassified as non-stroke.

This model showed a recall of 91.43% for No Stroke class and 96.77% for Stroke class, which represented good sensitivity of the model to the detection of stroke cases. The precisions were 96.97% and 90.91% for No Stroke and Stroke classes respectively. The performance was 93.94% and 6.06% misclassification rate, respectively, for the overall classification accuracy. The low false negative rate of 3.23% to diagnose stroke is of particular importance in clinical use, since a high false negative rate can lead to the underdiagnosis of true stroke cases. The results show that CNN model can balance and reliable predictions of brain stroke, and can handle well in distinguishing stroke and non-stroke CT images.

Classification Performance

The classification accuracy of the CNN and Hybrid NeuroResUNet models was assessed by accuracy, precision, recall, F1-score, specificity and the confusion matrix analysis. The CNN model was able to correctly classify 124 out of 132 samples with an overall accuracy of 93.94%. It has also shown high sensitivity (96.77%) in detecting stroke and a specificity of 91.43% indicating that it is well balanced in both classes and has less false negative rate.

The overall accuracy of the Hybrid NeuroResUNet model was 95.06% with a high degree of bias towards the majority class. It correctly identified most non-stroke cases but its recall to detect stroke cases was very low, with a high number of missed stroke cases. Although the overall accuracy of CNN is slightly lower than that of ResNet, its ability to identify stroke early and accurately is more reliable and clinically useful, which is important in clinical practice. Overall accuracy is not the only metric that should be used for assessing the medical diagnostic models, as the findings in these results suggest that class-wise performance metrics will be more useful.

Table 4. Comparative Analysis of Hybrid Model and CNN

Metric	Hybrid NeuroResUNet	CNN
Accuracy (%)	95.06	93.94
Precision (%)	100.00	90.91
Recall / Sensitivity (%)	5.00	96.77
Specificity (%)	100.00	91.43
F1-Score (%)	9.52	93.75
False Negative Rate (%)	95.00	3.23
Clinical Reliability	Low	High

The comparison between the Hybrid NeuroResUNet and CNN model with respect to brain stroke prediction is made in Table 4. The Hybrid NeuroResUNet had a slightly higher overall accuracy, however, it had an extremely low sensitivity towards stroke cases, which means that it had a high false negative rate. On the other hand, CNN model showed good recall, specificity, and F1-score with balanced performance, allowing for reliable detection of stroke and non-stroke cases. Although the CNN model has slightly lower overall accuracy, its superior ability to detect the stroke makes it more practical in clinical settings, where the goal is to minimize the missed diagnosis of stroke.

Error Analysis

The confusion matrix was used to compare the predictions made by the CNN model with the ground truth. The confusion matrix showed that the CNN model had an overall error rate of 6.06%, making 8 mistakes out of a total of 132 samples. In particular, six non-stroke cases were wrongly identified as stroke, and two stroke cases were wrongly identified as non-stroke cases. The false positive predictions can be explained by the fact that there are some common features in the images obtained from both normal and stroke patients, whereas the false negative can be explained by the fact that there might be some subtle features that are hard to distinguish in stroke images. In spite of the errors, the model remained with a high sensitivity (96.77%) which showed a good detection of stroke cases. This has a low false negative rate of 3.23% which is significant in clinical applications because of the risk of missed diagnosis of stroke. The overall results illustrate the capability of the CNN model to achieve competitive performance, with potential for further improvement with increased training data and better feature extraction methods.

Comparative results have shown that Hybrid NeuroResUNet edges out on overall accuracy but struggles in detecting critical strokes due to class imbalance, yielding lower recall for the minority class. In contrast, the CNN model has only slightly lower overall accuracy but enjoys a rather well-balanced performance, improving clinical utility by reliably highlighting most stroke cases. This represents a trade-off whereby medical AI should not be evaluated solely based on aggregated accuracy, but with particular emphasis on recall and F1-score with respect to clinically critical classes, such as stroke.

Class balancing, oversampling of the minority stroke class, cost sensitive learning, and optimization of the threshold value for the Hybrid NeuroResUNet model could improve the model in future works to minimize false negative predictions. Furthermore, the model could be enhanced by the addition of attention mechanisms or ensemble learning methods to further boost its discriminative power. Additionally, hyperparameter tuning and improved regularization techniques could be used to further optimize the CNN model for better performance and robustness.

In general, the Hybrid NeuroResUNet model performed slightly better in terms of accuracy, but in terms of classification performance, the CNN model had a more balanced performance and exhibited higher stroke detection sensitivity and clinical reliability. Thus, the CNN model is a more reliable and practical approach for automated brain stroke diagnosis, where the early diagnosis and accurate prediction of brain stroke is crucial.

5. CONCLUSION

In this study, the performance of a CNN and Hybrid NeuroResUNet model were compared with respect to automated brain stroke prediction using CT scan images. Experimental results demonstrated that the Hybrid NeuroResUNet had a relatively high overall accuracy of 95.07%, however, the model performance was still low because of the high value of false negative rate, thus low sensitivity to stroke cases. This limitation will make it less effective for clinical use, where timely and precise stroke detection is important.

The CNN model, on the other hand, exhibited the best and balanced performance, with an accuracy of 93.94%, a sensitivity of 96.77%, a specificity of 91.43%, and an F1 score of 93.75%. The confusion matrix analysis also showed that the CNN approach was successful in accurately diagnosing most stroke and non-stroke patients and had a low misclassification rate. The results indicate that the CNN model outperformed the Hybrid NeuroResUNet in terms of clinical reliability and balanced classification performance. To further optimize the detection of strokes and minimize false negative predictions, further work will involve optimizing the Hyperparameters of the Hybrid NeuroResUNet model, applying Attention Mechanism and Class Balancing techniques.

Data Available

In this work, we used a dataset of 5185 CT scan images obtained from Kaggle, which is recognized for its quality, accessibility, and relevance to the research aims. Brain Stroke CT Image Dataset. [Online]. Available: <https://www.kaggle.com/datasets/afriDirahman/brain-stroke-ctimage-dataset>.

References:

1. Ramya Sree, K., MohanKumar, P., "A Comparative Analysis of Prediction of Brain Stroke Using AIML" In: Abraham, A., Pllana, S., Casalino, G., Ma, K., Bajaj, A. (eds) Intelligent Systems Design and Applications. ISDA 2022. Lecture Notes in Networks and Systems, vol 715. Springer, Cham. https://doi.org/10.1007/978-3-031-35507-3_52
2. Carlo Domenico Maida, Rosario Luca Norrito, Salvatore Rizzica, Marco Mazzola, Elisa Rita Scarantino, Antonino Tuttolomondo, "Molecular Pathogenesis of Ischemic and Hemorrhagic Strokes: Background and Therapeutic Approaches," International Journal of Molecular Sciences 2024, 25(12), 6297; <https://doi.org/10.3390/ijms25126297>
3. Li X, Chen G., "CNS-peripheral immune interactions in hemorrhagic stroke," Journal of Cerebral Blood Flow and Metabolism.2023;43(2):185-197. doi:10.1177/0271678X221145089, <https://journals.sagepub.com/doi/10.1177/0271678X221145089>
4. Enrico Fainardi, Giorgio Busto, Andrea Morotti, "Automated advanced imaging in acute ischemic stroke: Certainties and uncertainties" European Journal of Radiology Open, Volume 11, 2023,100524, ISSN 2352-0477, <https://doi.org/10.1016/j.ejro.2023.100524>
5. Kalaria, R.N., Akinyemi, R., Ihara, M., "Stroke injury, cognitive impairment, and vascular dementia. Biochim," Biophys. Acta 1862, 915–925, 2016, <https://doi.org/10.1016/j.bbadis.2016.01.015>
6. Genevieve Richarda, Knut Kolskar, Kristine M. Ulrichsen, Tobias Kaufmann, Dag Alnaes, Anne-Marthe Sanders, Erlend S. Dorum, Jennifer Monereo Sánchez, Anders Petersen, Hege Ihle-Hansen, Jan Egil Nordvik, Lars T. Westlye, "Brain age prediction in stroke patients: Highly reliable but limited sensitivity to cognitive performance and response to cognitive training", NeuroImage: Clinical 25 (2020) 102159, <https://doi.org/10.1016/j.nicl.2019.102159>

7. Eman M Alanazi, Aalaa Abdou, Jake Luo, "Predicting Risk of Stroke from Lab Tests Using Machine Learning Algorithms: Development and Evaluation of Prediction Models," *JMIR Formative Research*, 2021, vol. 5, iss. 12: e23440, doi: 10.2196/23440
8. Nitish Biswas, Khandaker Mohammad Mohi Uddin, Sarreha Tasmin Rikta, Samrat Kumar Dey, "A comparative analysis of machine learning classifiers for stroke prediction: A predictive analytics approach," *Healthcare Analytics 2* (2022), <https://doi.org/10.1016/j.health.2022.100116>
9. Vamsi Bandi, Debnath Bhattacharyya, Divya Midhun chakkravarthy, "Prediction of Brain Stroke Severity Using Machine Learning", *Revue d'Intelligence Artificielle*, Vol. 34, No. 6, December 2020, pp. 753-761, <https://doi.org/10.18280/ria.340609>
10. M. M. Islam, S. Akter, M. Rokunojjaman, J. H. Rony, A. Al Amin, and S. Kar. 2021. Stroke Prediction Analysis using Machine Learning Classifiers and Feature Technique. *International Journal of Electronics and Communications System*, vol. 1, no. 2, 17-22, 2021. doi: 10.24042/ijecs. v1i2.10393, <https://ejournal.radenintan.ac.id/index.php/IJECS/article/view/10393/pdf>
11. Samaa A. Mostafa, Doaa S. Elzanfaly, Ahmed E. Yakoub., "A Machine Learning Ensemble Classifier for Prediction of Brain Strokes," *International Journal of Advanced Computer Science and Applications*, Vol. 13, No. 12, 2022. doi: 10.14569/IJACSA.2022.0131232, https://www.researchgate.net/publication/366757731_A_Machine_Learning_Ensemble_Classifier_for_Prediction_of_Brain_Strokes
12. Rukiye Karakis, Kali Gurkahraman, Georgios D. Mitsis, Marie-Helene Boudrias, "Deep learning prediction of motor performance in stroke individuals using neuroimaging data," *Journal of Biomedical Informatics* 141 (2023) 104357, <https://doi.org/10.1016/j.jbi.2023.104357>
13. A. Srinivas a, b, Joseph Prakash Mosiganti, "A brain stroke detection model using soft voting-based ensemble machine learning classifier," *Measurement: Sensors* 29 (2023) 100871, <https://doi.org/10.1016/j.measen.2023.100871>
14. Muhammad Asim Saleem, Ashir Javeed, Wasan Akarathanawat, Aurauma Chutine, Nijasri Charnnarong Suwanwela, Widhyakorn Asdornwised, Surachai Chaitusaney, Sunchai Deelertpaiboon, Wattanasak Srisiri, Watit Benjapolakul, Pasu Kaewplung., "Innovations in Stroke Identification: A Machine Learning-Based Diagnostic Model Using Neuroimages," *IEEE Access*, 2024, vol. 12, pp. 35754-35764, 2024, doi: 10.1109/ACCESS.2024.3369673. <https://ieeexplore.ieee.org/document/10445193>
15. Brain Stroke CT Image Dataset. Accessed: Jul. 3, 2023. [Online]. Available: <https://www.kaggle.com/datasets/afdirahman/brain-stroke-ctimage-dataset>
16. A. Hingmire, A. Sheikh, S. Shukla, R. Shah, S. Jadhav, T. P. Nagarhalli, and A. Ruperee, "Intelligent Crime Scene Recognition: Advancing Public Safety through Deep Learning Architectures and Event Sequence Analysis," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 11, no. 4, pp. 64–71, Apr. 2024, <https://doi.org/10.14445/23488379/IJEEE-V11I4P108>
17. H. Jiang and J. Zhao, "Multi-lesion Segmentation of Fundus Images using Improved UNet++," *IAENG International Journal of Computer Science*, vol. 51, no. 10, pp. 1587–1595, Oct. 2024.
18. Y. G. Wang, L. Wang and Y. X. Geng, "Research on Medical Image Classification Based on Triple Fusion Attention", *IAENG International Engineering Letters*, vol. 33, no. 1, pp. 124–131, January 2025, https://www.iaeng.org/IJCS/issues_v51/issue_10/IJCS_51_10_16.pdf
19. Gangavarapu Sailasya and Gorli L Aruna Kumari. "Analyzing the Performance of Stroke Prediction using ML Classification Algorithms". *International Journal of Advanced Computer Science and Applications* (IJACSA) 12.6 (2021). <http://dx.doi.org/10.14569/IJACSA.2021.0120662>
20. Aniwat Phaphuangwittayakul, A., Guo, Y., Ying, F. et al. An optimal deep learning framework for multi-type hemorrhagic lesions detection and quantification in head CT images for traumatic brain injury. *Appl Intell* 52, 7320–7338 (2022). <https://doi.org/10.1007/s10489-021-02782-9>
21. B. Akter, A. Rajbongshi, S. Sazzad, R. Shakil, J. Biswas and U. Sara, "A Machine Learning Approach to Detect the Brain Stroke Disease," 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2022, pp. 897-901, <https://doi.org/10.1109/ICSSIT53264.2022.9716345>
22. Azhar Tursynova, Bатыркhan Omarov, Natalya Tukenova, Indira Salgozha, Onergul Khaaval, Rinat Ramazanov, Bagdat Ospanov. "Deep Learning-Enabled Brain Stroke Classification on Computed Tomography Images", *Computers, Materials & Continua* 2023, 75(1), 1431-1446. <https://doi.org/10.32604/cmc.2023.034400>
23. S. Xun, Y. Zhang, S. Duan, M. Wang, J. Chen, T. Tong, Q. Gao, C. Lam, M. Hu, and T. Tan, "ARGA-Unet: Advanced Unet segmentation model using residual grouped convolution and attention mechanism for brain tumor MRI image segmentation," *Virtual Reality & Intelligent Hardware*, vol. 6, no. 3, pp. 203–216, 2024. doi: 10.1016/j.vrih.2023.05.001.
24. Aakanksha Gupta, Nidhi Mishra, Nishtha Jatana, Shaily Malik, Khaled A. Gepreel, Farwa Asmat, Sachi Nandan Mohanty, "Predicting stroke risk: An effective stroke prediction model based on neural networks", *Journal of Neurorestoratology*, Volume 13, Issue 1, February 2025, 100156, <https://doi.org/10.1016/j.jnrt.2024.100156>
25. Y. Sailaja and P. Velmurugan, "A Novel Feature Extraction and Deep Spiking MobileNet for Brain Stroke Detection Using MRI Image," *Engineering Applications of Artificial Intelligence*, vol. 159, Part A, Art. no. 111565, Nov. 2025. doi: 10.1016/j.engappai.2025.111565.
26. Mahesh Anil Inamdar, Anjan Gudigar , U. Raghavendra, Massimo Salvi, Nithin Raj, J. Pooja, Ajay Hegde, Girish R. Menon, U. Rajendra Acharya, "Dual attention mechanisms with patch-level significance embedding for ischemic stroke

- classification in brain CT images”, *Informatics in Medicine Unlocked*, 57 (2025) 101678, <https://doi.org/10.1016/j.imu.2025.101678E>.
27. Mostafapour, M. Memarian, Ebrahimi, N. Ezzati, and A. Aminfar, “Brain Stroke Detection from MRI Images Using Convolutional Neural Networks and XGBoost,” *Biomedical Signal Processing and Control*, vol. 113, Part A, Art. no. 109049, Mar. 2026. doi: 10.1016/j.bspc.2025.109049.
 28. M. Shakunthala, T. D. Subha, A. Anna Lakshmi, and T. D. Subash, “Automated Brain Stroke Detection in Magnetic Resonance Imaging Using Self-Modulating Convolutional Neural Networks for Enhanced Ischemic and Hemorrhagic Stroke Classification,” *Knowledge-Based Systems*, vol. 341, 115665, May 2026. doi:10.1016/j.knosys.2026.115665