AN EXPLAINABLE AI DRIVEN BRAIN STROKE DETECTION: A DEEP LEARNING FRAMEWORK USING CT SCAN IMAGES

Shaik Amesha Muskan
PG Scholar
Department of Computer Science & Engineering
JNTUA College of Engineering (Autonomous)
Ananthapuramu, Andhra Pradesh, India
ameshamuskan9@gmail.com

Dr. R. Raja Sekhar
Professor
Department of Computer Science & Engineering
JNTUA College of Engineering (Autonomous)
Ananthapuramu, Andhra Pradesh, India
drasharaj2002.cse@jntua.ac.in

Abstract: *Stroke is a potentially fatal illness that requires rapid diagnosis quickly in order to prevent serious complications. This study suggests use the brain CT scan images to create an automated stroke detection system combined with explainable artificial intelligence (XAI). A dataset consisting of normal and stroke images was pre-processed, and data augmentation was performed to improve model performance and address class imbalance. Multiple deep learning models, including CNN, ResNet18, MobileNetV3 and a custom Radiological ImageNet, were trained and evaluated under different experimental settings, such as original and augmented data. However, only augmentation was applied to CNN and RadImageNet. Among these, ResNet18 achieved the highest accuracy than other models because of its deep neural networks and using of skip connection. The performance of the model was evaluated using various metrics and was validated against new data to confirm its ability to generalize. Grad-CAM was applied to generate heatmaps for the visual interpretation of model predictions. The system was further deployed as a dashboard for real-time analysis, providing an efficient and interpretable solution to assist clinicians in stroke diagnosis.*

Keywords: Brain Stroke, CT scans, Deep learning, ResNet18, MobileNeV3, CNN, Radiological ImageNet Explainable AI, Grad-CAM.

1. INTRODUCTION:

Stroke ranks among the top causes of mortality and prolonged disability across the globe, necessitating quick and precise diagnosis for successful treatment. Brain scan images especially Computed Tomography (CT), are commonly employed in clinical settings because of their rapidity, accessibility, and efficiency in identifying brain abnormalities.

However, manual interpretation of CT images is challenging and highly dependent on radiological expertise, which may lead to variability and delayed diagnosis, especially in early-stage stroke cases. stroke detection using CT imaging [1], [2], [3].

In recent years, deep learning techniques have gained significant attention in medical image analysis due to their ability to automatically extract complex features from imaging data. Convolutional Neural Networks (CNNs) have shown strong performance in classification tasks, including stroke detection from CT images [4], [5]. Advanced architectures such as ResNet18 improve model performance through residual learning, enabling deeper networks to learn effectively [6]. Additionally, hybrid and transformer-based approaches further enhance classification accuracy in complex medical datasets [7], [8].

Along with accuracy, computational efficiency is essential for real-time medical applications. Lightweight architectures such as MobileNetV3 and other efficient deep learning frameworks reduce computational cost while maintaining high performance, making them suitable for practical deployment in healthcare systems. These models are particularly useful when handling large-scale medical imaging data with limited computational resources.

Despite these advancements, the lack of interpretability remains a major challenge in applying deep learning models in clinical settings. Explainable Artificial Intelligence (XAI) techniques, such as Grad-CAM, provide visual explanations by highlighting important regions in CT images that influence model predictions [9], [10].

In this study, we propose an Explainable AI-driven deep learning framework for brain stroke detection using CT scan images. The system incorporates preprocessing and data augmentation techniques to enhance model robustness and address class imbalance. Multiple models, including CNN, ResNet18, MobileNetV3, and a custom RadNet, are trained and evaluated under different experimental conditions. Furthermore, Grad-CAM is integrated to provide visual explanations of predictions, improving model transparency. The proposed System is designed to provide both accuracy, efficiency and interpretable solution to assist clinicians in early stroke diagnosis and decision-making

2. RELATED WORKS

Recent progress in machine learning and deep learning has greatly enhanced the identification and diagnosis of brain strokes through the use of medical imaging and clinical information. Numerous research efforts have investigated diverse methods, from conventional machine learning models to sophisticated deep learning frameworks and interpretable systems.

Research on stroke identification utilizing transfer learning showed the efficacy of deep learning architectures like CNN, ResNet, MobileNet, and RadiNet in categorizing brain CT scan images [11]. The writers utilized data enhancement methods to address class imbalance and enhance model performance. Of the assessed models, ResNet18 delivered outstanding outcomes with excellent accuracy, precision, and recall, underscoring the significance of lightweight architectures in analyzing medical images. This study highlights that transfer learning can greatly improve stroke detection efficacy while minimizing computational complexity.

A web-based stroke risk assessment system utilizing machine learning methods applied to patient medical records was suggested in a different study [12]. The system used logistic regression for prediction and SMOTE (Synthetic Minority Oversampling Technique) to handle class imbalance, with a 93.2% accuracy rate. Because the model was developed using an intuitive interface, the primary focus of this research is on making it interpretable and accessible. However, risk prediction is the main focus of this method rather than image-based diagnosis.

Additionally, a sophisticated neuroimage-based machine learning diagnostic system that combines feature selection with deep learning architectures like Bidirectional Long Short-Term Memory (BiLSTM) networks was shown [13]. With an accuracy of 96.5%, the suggested approach extracted the best features from CT scans using genetic algorithms. The study showed that feature selection techniques and deep learning models can improve classification accuracy and yield more trustworthy diagnostic outcomes. However, these hybrid models' computational complexity remains an issue.

Using EfficientNet, GhostNet, and optimization techniques, a novel deep learning architecture for stroke diagnosis utilizing MRI images was suggested in another recent study [14]. By collecting pertinent features and improving them using cutting-edge methods, the model attained excellent accuracy. This study demonstrates how deep learning may be used to classify strokes with extreme accuracy. However, it also highlights difficulties including the need for sizable annotated datasets and problems with generalization under various imaging settings.

In the same way, Sakri et al. focused on both classification performance and interpretability when creating an enhanced concatenation of deep learning models for ischemic stroke prediction and interpretation. In order to facilitate clinical decision-making, their work emphasized the significance of combining prediction accuracy with model interpretation [15]. Furthermore, only a small number of research use Explainable AI (XAI) strategies to increase the model's transparency.

In order to fill these gaps, the proposed work focuses on creating an Explainable AI-driven deep learning framework that leverages methods like Grad-CAM to provide visual explanations in addition to achieving high accuracy in stroke identification using CT scan pictures. The system is more dependable and appropriate for clinical applications because of its performance and interpretability.

3. PROBLEM STATEMENT

Stroke necessitates prompt and precise diagnosis to avert severe complications; however, manual analysis of CT scan images is challenging and reliant on the proficiency of radiologists. Signs of a stroke in its early stages are often hard to spot, which can lead to a wrong diagnosis or a delay in treatment. Deep learning models can make detection more accurate, but there are still problems like class imbalance, overfitting, and a lack of interpretability. Most models are black boxes, which makes it hard to trust their decisions in medical settings. So, there needs to be an automated system that can accurately find strokes in CT images and give clear visual explanations to help doctors make good decisions.

EXISTING SYSTEM:

In the Existing System, the stroke detection is based upon predicting normal and stroke using CT scan images but lack of interpretability and it uses multiple models and algorithms for classification this leads to heavy computational cost and there is no explainability for images.

4.PROPOSED METHODOLOGY:

The methodology for "An Explainable AI-driven brain stroke detection: A deep learning framework using CT scan images" project using the different deep convolutional neural networks and pretrained modelsmused to classify stroke vs normal. For our binary classification challenge of predicting brain stroke using CT images, we assessed and contrasted four different designs from CNN, MobileNetV3, ResNet18 and Radiological ImageNet and selected the best-performing architecture. To expand the dataset, data augmentation and image preprocessing are carried out. Following implementation, we test the model and determine whether a CT scan indicates a stroke or not.

DATASET COLLECTION:

The Brain Stroke CT Image Dataset is accessible online. Brain CT scans are from the Kaggle website. There are 950 scans of grayscale, JPG-formatted scans that are classified as having a brain stroke and 1551 that are classified as normal. This dataset, which has 2501 images total, follows an uneven class size and imbalance in the quantity of images representing cases with and without strokes. In this dataset, every image has a label as "Normal" or "Stroke" in order to conduct research and analysis.

DATA PREPROCESSING AND SPLITTING:

Data preprocessing was performed to standardize CT scan images and improve model performance. To guarantee consistent input and reliable training, all images were normalized and scaled to a fixed dimension of 224×224. CT scans were processed in accordance with the model's specifications because they are grayscale. The data is divided into train and test datasets which consists of total images 2501 and [train, test] = [2000,501].

DATA AUGMENTATION:

To improve the dataset's diversity and reduce overfitting, data augmentation techniques like rotation, flipping, and scaling were incorporated to weaker and simpler models like CNN, Radiological ImageNet. These preprocessing steps ensure the accuracy of the stroke recognition, feature learning, and generalization.

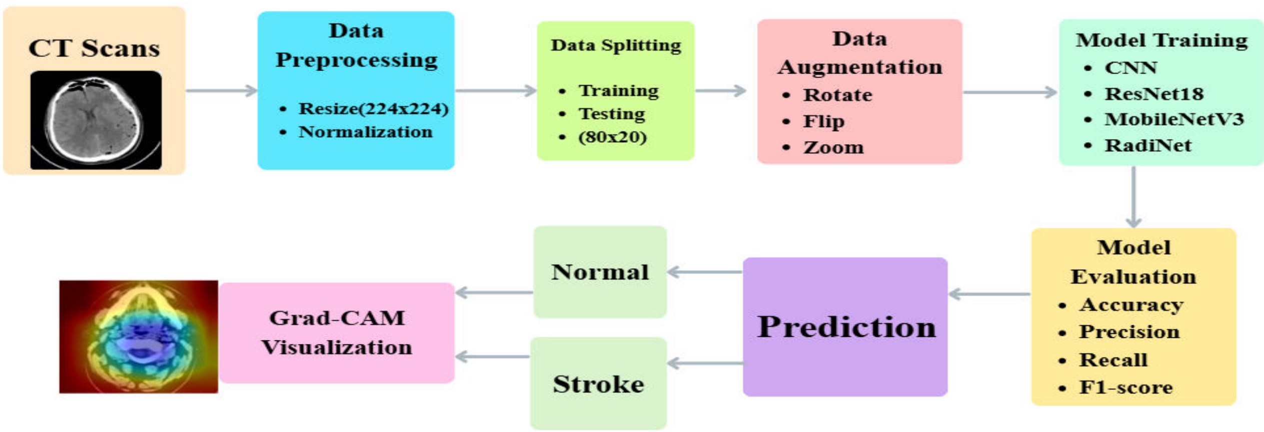

Fig.1 Proposed System Workflow

MODEL TRAINING:

This includes four deep CNN models such as, MobileNetV3, ResNet18 and Radiological ImageNet & selected the best-performing architecture.

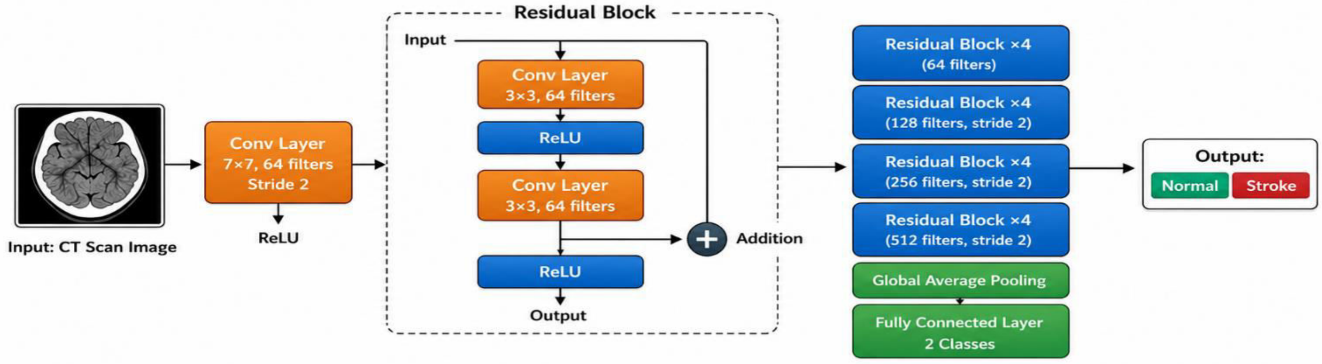

Fig.2 ResNet18 Architecture

4.1. CNN (Convolutional Neural networks):

Convolutional Neural Networks (CNNs), which are specific types of deep learning models, are used in the processing of visual data and other associated operations like object identification and image classification. They are considered the primary models in the extraction of features. CNN uses convolution layers to extract features from CT images. These layers contain filters that help in detecting edges, textures, and intensity. ReLU is used for non-linearity, and pooling is used for reducing dimensions. These layers are used to learn complex features, such as stroke regions. Finally, fully connected layers classify images as normal or stroke. Our advanced stroke recognition system uses Convolutional Neural Networks (CNNs), which are specific types of deep learning models

4.2. MobileNetV3:

When it comes to deep learning models, the MobileNetV3 model's lightweight structure proven to be quite efficient in the process of information extraction from the CT scan of the brain. In MobileNetV3, depth-wise separable convolution is utilized to minimize the computational cost without compromising the trained model accuracy. In the project, the bottleneck and attention blocks of the model enabled the efficient extraction of features related to the strokes and improved the accuracy of the model in comparison with other models. The model is using the concept of transfer learning with the ImageNet dataset and fine-tuning the model on the stroke dataset for accurate classification. The MobileNetV3 model is quite efficient in the process of capturing the spatial and structural features of the stroke region and accurately classifying the image using the concept of fully connected layers

The actual training code of a pre-trained MobileNetV3-Large model was utilized for brain stroke classification using CT scan images. The final classification layer was modified to produce two output classes: Normal and Stroke. The model was trained for 10 epochs using the Adam optimizer with a learning rate of 0.001 and Cross-Entropy Loss as the objective function. During training, forward propagation, loss computation, backpropagation, and parameter optimization were performed iteratively. Validation was conducted after each epoch to evaluate model generalization, and the best-performing model was saved based on validation accuracy.

4.3. ResNet18:

In ResNet18, residual blocks with skip connections are employed. In this case, the original image is added directly to the output of the convolutional layers. This ensures the retention of original CT image information and the acquisition of new information. The model uses the residual learning method to learn complex features by eliminating the problem of vanishing gradients. In the project, the ResNet18 model has been fine-tuned using the pre-trained weights and utilized for the classification of images as normal or stroke classes. In the case of stroke detection, the model can capture the overall brain structure and abnormalities. However, the ResNet18 model has outperformed than models with 99% accuracy. The residual block follows the formula where; the input is added to the output of convolution layers. This residual learning helps prevent vanishing gradients, enables deeper network training, and improves feature extraction for accurate brain stroke classification from CT scan images

$$y = F(x) + x$$

However, the training code explains ResNet18 was trained using transfer learning by replacing its final classification layer with a two-class output layer. During training, CT images were passed through the network, Cross-Entropy Loss was computed, and weights were updated using the Adam optimizer. After each epoch, validation accuracy was measured, and the best-performing model was saved.

4.4. Radiological ImageNet (RadImageNet):

Radiological ImageNet is a customized CNN model developed to perform image analysis on CT images. It consists of convolution, activation, and pooling layers to detect image features like edges, textures, and abnormalities. Radiological ImageNet is simpler compared to the standard CNN model; however, it is computationally efficient. It can be applied to small image datasets. Radiological ImageNet can assist in learning stroke-specific image features in our project. It is a customized CNN designed to extract features specific to CT images. The architecture of Radiological ImageNet is similar to CNN model.

After training is complete, we take out the best model based on accuracy and predictions. The Graph is generated based upon the high and low accuracy.

5.MODEL EVALUATION:

Using key metrics like accuracy, precision, recall, and F1-score to assess the trained models' performance. This stage provides accurate stroke detection and helps in identifying the optimal model

6.MODEL PREDICTION:

Using trained model to determine if a certain brain CT scan is indicative of a stroke or a normal scan. The model will provide real-time predictions along with Grad-CAM visualizations to explain the decision-making process.

7. MODEL EXPLAINABILITY USING GRAD-CAM

Grad-CAM Visualization

For better interpretation of deep learning models, the approach of Grad-CAM was implemented to visualize the part of brain CT scan images that contributed the most towards making the classification decision. Once the model classified an image as Normal or Stroke it will produce output in the form of heatmaps.

In the proposed brain stroke detection system, Grad-CAM serves as a visual explanation for the decisions made by the model by highlighting the important brain parts. Areas shown in warmer colors (red and yellow) indicate regions with greater importance, while cooler colors (blue and green) show lower relevance. This visualization method boosts model transparency, helps clinicians understand the prediction process, and builds trust in the automated diagnostic system.

8. RESULTS AND DISCUSSION

8.1 Experimental Results:

The proposed system was tested with CNN, ResNet18, MobileNetV3, and RadNet under different conditions of the dataset. The highest accuracy was obtained by the ResNet18 model at 99%. The second-best accuracy was obtained by the MobileNetV3 model at 96%.

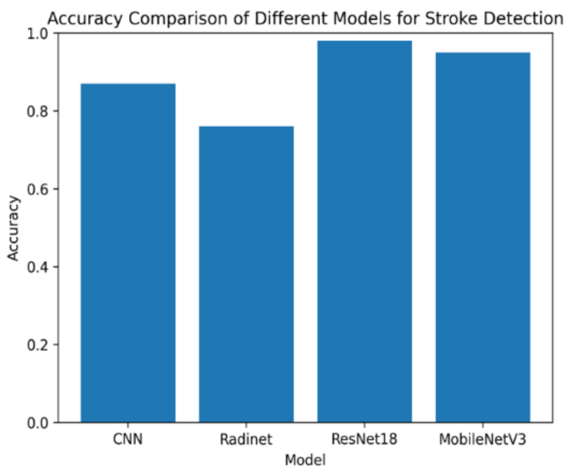

Fig.3 Bar Graph of all models

CNN Model:

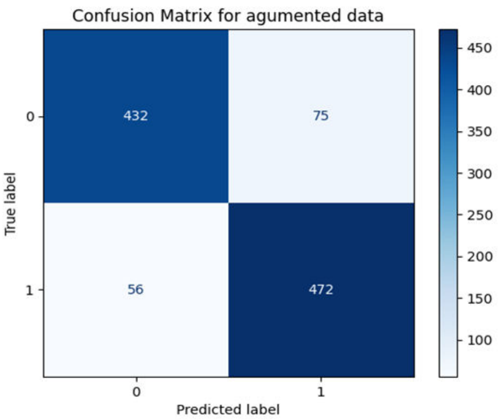

Fig.4 Confusion Matrix of CNN

MobileNetV3:

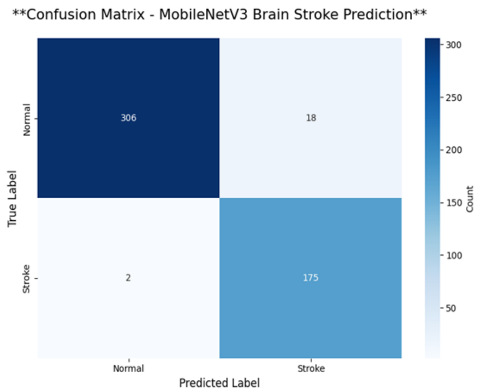

Fig.5 Confusion Matrix of MobileNetV3

ResNet18:

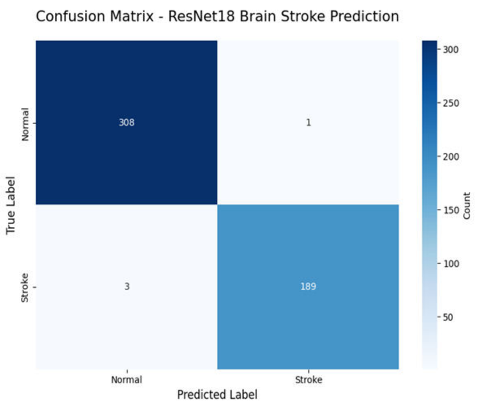

Fig.6 Confusion Matrix of ResNet18

RadImageNet:

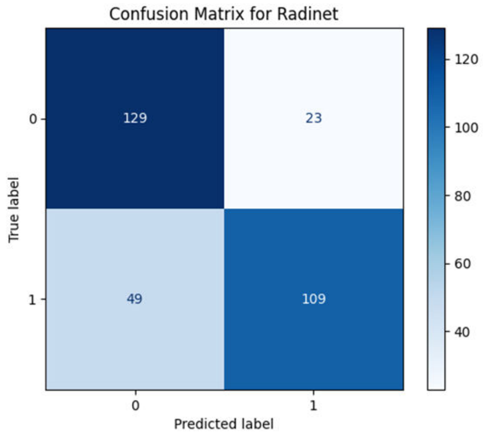

Fig.7 Confusion Matrix of RadImageNet

8.2 COMPARATIVE ANALYSIS:

Model	Accuracy	Precision	Recall	F1-score
CNN	0.87	0.86	0.89	0.88
ResNet18	0.99	0.99	0.98	0.99
MobileNetV3	0.96	0.91	0.99	0.95
RadImageNet	0.77	0.83	0.69	0.75

8.3 EXPLAINABILITY ANALYSIS:

For improved interpretability, Grad-CAM (gradient class activation mapping) displayed the regions of the CT scans used in the model's predictions. From the heat maps, it is evident that areas of the brain that were affected, especially in stroke, were highlighted. In normal cases, the highlighted region is defined as an important area rather than the stroke itself, since the model's decision is based on its training (learned features). This implies that the model is able to focus on relevant areas of the images, excluding irrelevant information in the background.

This is a very important aspect of this component, especially in medical imaging, where a high level of validation is required. This ensures that, in addition to being accurate, the framework is reliable.

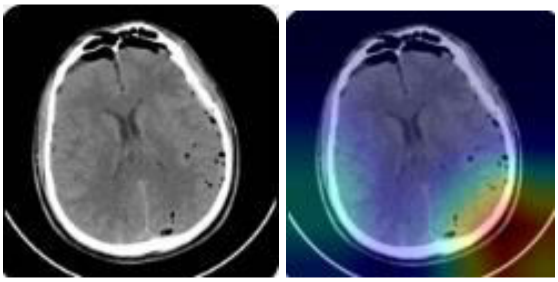

Fig:8 Prediction (Normal)

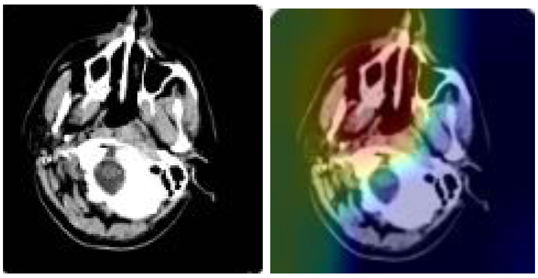

Fig:9 Prediction (Stroke)

8.4 DISCUSSIONS:

The results indicate that the deep learning models, especially the ResNet18 model, are highly effective in the detection of brain stroke by analyzing the images obtained through the CT scan technique.

First and foremost, it has been observed that even the lightweight models may perform better than the complex ones by tuning the models appropriately, especially in cases where the dataset may be limited or imbalanced. Secondly, the addition of the explainability technique has removed the major limitation of the deep learning-based models.

However, the study has some limitations as well. For instance, the dataset used in the study may not be sufficient, and the results may vary in the case of a large dataset.

9. CONCLUSION

The paper introduced an Explainable AI-driven framework based on a deep learning model for automated detection of brain stroke using images obtained from a CT scan.

Various models were developed and tested under various experimental conditions. Out of all models, including CNN, MobileNetV3, ResNet18 and RadImageNet, ResNet18 obtained the highest accuracy. This shows that this model is quite efficient in extracting meaningful features. The incorporation of Explainable Artificial Intelligence using Grad-CAM also helped in visualizing the regions that are responsible for model predictions. This shows that this framework is quite efficient in terms of interpretability. This framework can be used for real-time diagnosis using a dashboard. This system is helpful in creating efficient AI models for medical image analysis.

The future work can be focused on the development of the proposed system for multi-class classification of strokes, i.e., ischemic and haemorrhagic strokes. In addition, the system can be trained on more diverse datasets. Moreover, the usage of other powerful architectures like Vision Transformers and other XAI tools other than Grad-CAM can be explored for the improvement of the system's performance. Additionally, the usage of other modalities like MRI images and the optimization of the proposed system for edge devices can be explored for the improvement of the system's applicability.

References

[1] R. Karthik et al., "Neuroimaging and Deep Learning for Brain Stroke Detection: A Review of Recent Advancements," Comput. Methods Programs Biomed., vol. 197, 2020.

[2] R. Halder and N. Sharmin, "Brain Ischemic Stroke Detection Through Deep Learning: A Systematic Review on CT vs MRI," IJEME, 2024.

[3] J. A. Bojsen et al., "Artificial Intelligence for Stroke Detection: A Systematic Review and Meta-Analysis," Insights into Imaging, 2024.

[4] H. Abdi et al., "Stroke Detection in Brain CT Images Using Convolutional Neural Networks: Model Development and Interpretability," Information, vol. 16, no. 5, 2025.

[5] M. S. Hossen et al., "An Efficient Deep Learning Framework for Brain Stroke Diagnosis Using CT Images," arXiv, 2025.

[6] D. Litjens et al., "A Survey on Deep Learning in Medical Image

Analysis," Med. Image Anal., 2017. (Background support)

[7] S. Qari and M. A. Thafar, "Brain Stroke Classification Using CT Scans with Transformer-Based Models and Explainable AI," Diagnostics, 2025.

[8] StrokeNeXt Study, "An Automated Stroke Classification Model Using CT and MRI Images," BMC Medical Imaging, 2025.

[9] B. Ignac et al., "Deep Learning Model with Grad-CAM for CT Image Analysis," Applied Sciences, 2026.

[10] S. Selvaraju et al., "Grad-CAM: Visual Explanations from Deep Networks," ICCV, 2017.

[11] A. Anand et al., Identifying Brain Strokes by Image Analysis Using Transfer Learning, 2025.

[12] D. Bhandari et al., A Web-Based Interface That Leverages Machine Learning to Assess an Individual's Vulnerability to Brain Stroke, IEEE Access, 2025.

[13] M. A. Saleem, A. Javeed, W. Akarathanawat, A. Chutinet, N. C. Suwanwela, and W. Asdornwised, ''Innovations in Stroke Identification: A Machine Learning-based Diagnostic Model Using Neuroimages'' IEEE Access, vol. 12, pp. 35754–35764, 2024

[14] X. Xue, S. E. Viswapriya, D. Rajeswari, R. Z. Homod, and O. I. Khalaf, ''An efficient deep learning network for brain stroke detection using salp shuffled shepherded optimization" Scientific Reports, vol. 15, pp. 33516, 2025.

[15] S. Sakri et al., "An Improved Concatenation of Deep Learning Models for Predicting and Interpreting Ischemic Stroke," *IEEE Access*, vol. 12, pp. 53189–53204, 2024.