Brain Stroke Diagnosis Using Auxiliary Branch Guided Swin Transformer with Pseudo-Segmentation Supervision

Batyrkhan Omarov
Narxoz University, Kazakhstan | International Information Technology University, Kazakhstan | Kh. Dosmukhamedov Atyrau State University, Kazakhstan
batyahan@gmail.com (corresponding author)

Zhanseri Ikram
Narxoz University, Kazakhstan
zhanserikaz@gmail.com

Received: 8 June 2025 | Revised: 22 July 2025 | Accepted: 29 July 2025

 DOI: https://doi.org/10.48084/etasr.12602

ABSTRACT

Accurate and timely diagnosis of brain stroke is critical for effective clinical intervention and long-term patient outcomes. In this study, a novel deep learning-based framework for automated stroke diagnosis is proposed, utilizing an Auxiliary Branch Guided Swin Transformer with pseudo-segmentation supervision. The proposed architecture combines the hierarchical representation power of the Swin Transformer with a parallel auxiliary segmentation branch to enhance lesion-specific attention and spatial awareness. To address the scarcity of detailed annotations in clinical datasets, we employ pseudo-labels generated from bounding box–level supervision, enabling the model to learn lesion localization without full pixel-wise segmentation masks. The model was trained and validated on the ISLES 2024 dataset, which includes multimodal brain MRI scans. Quantitative results demonstrate that the proposed model achieves 94.6% accuracy, 94.3% precision, 94% recall, and an F1-score of 94%, outperforming existing CNN-based and transformer-based approaches. The auxiliary branch not only facilitates better feature refinement but also improves generalization by promoting regularization during training. This study highlights the effectiveness of transformer-based architectures in medical image analysis and introduces a practical solution for weakly-supervised stroke detection, offering a promising tool for clinical decision support and automated neuroimaging diagnostics.

Keywords-stroke diagnosis; swin transformer; auxiliary branch; pseudo-segmentation; brain MRI; deep learning; ISLES 2024

I. INTRODUCTION

The rapid and accurate diagnosis of brain stroke [1] type and location is critical for initiating timely therapeutic interventions, which directly influence clinical outcomes [2]. In recent years, neuroimaging modalities, particularly Diffusion-Weighted Imaging (DWI) and Perfusion-Weighted Imaging (PWI), have become the cornerstone for ischemic stroke diagnosis and tissue viability assessment [3]. However, the interpretation of such multimodal datasets is both time-consuming and highly dependent on expert radiological input, necessitating the development of intelligent, automated diagnostic tools. Artificial intelligence (AI) and, more specifically, Deep Learning (DL) have transformed the landscape of medical image analysis by enabling automatic feature extraction and lesion localization with remarkable accuracy [4]. Convolutional Neural Networks (CNNs) have demonstrated success in stroke lesion segmentation, but their limited receptive field and inability to model long-range dependencies present significant limitations [5]. To overcome these challenges, transformer-based architectures, such as the Swin Transformer, have emerged as promising alternatives by integrating hierarchical feature representation with shifted windows to capture contextual relationships at multiple scales [6]. Their ability to model global context makes them suitable for complex medical tasks like ischemic stroke diagnosis. While segmentation models have achieved notable performance, many require extensive pixel-wise annotations, which are scarce in real-world clinical datasets [7]. To address this, pseudo-segmentation supervision has been proposed as an effective surrogate strategy, using weak or generated labels to approximate segmentation maps during training [8]. Moreover, auxiliary branches in neural architectures have been shown to enhance learning by injecting intermediate supervision and encouraging more discriminative feature representations [9]. Integrating these concepts, this study proposes an auxiliary

branch guided Swin Transformer enhanced with pseudo-segmentation supervision for stroke diagnosis. The proposed model was trained and evaluated on the ISLES 2024 dataset [10], a benchmark for stroke lesion segmentation and outcome prediction using multimodal MRI. Our architecture aims to bridge the gap between clinical feasibility and computational precision by delivering accurate lesion localization without requiring exhaustive manual annotations. This study contributes to the growing field of AI-assisted neuroradiology by demonstrating how hybrid transformer frameworks can yield state-of-the-art performance in challenging diagnostic scenarios.

II. MATERIALS AND METHODS

This section outlines the experimental design, dataset characteristics, model architecture, and evaluation criteria employed in this study to develop and assess the proposed stroke diagnosis framework. The primary objective is to leverage an Auxiliary Branch Guided Swin Transformer architecture enhanced with pseudo-segmentation supervision to accurately classify and localize stroke lesions from brain MRI data. This section provides a comprehensive description of the neural network components, the preprocessing pipeline applied to the ISLES 2024 dataset, and the quantitative metrics used to evaluate performance. All methodological choices were guided by clinical relevance, reproducibility, and the need for robust generalization across diverse imaging scenarios.

A. Proposed Model Architecture

The architecture of the proposed model is illustrated in Figure 1. The input to the system is a single-channel axial slice of a brain MRI scan with dimensions B×1×224×224, where B denotes the batch size. The primary objective of the architecture is to perform three synergistic tasks: stroke classification, lesion localization, and pseudo-segmentation, under a multi-task learning framework guided by an auxiliary branch.

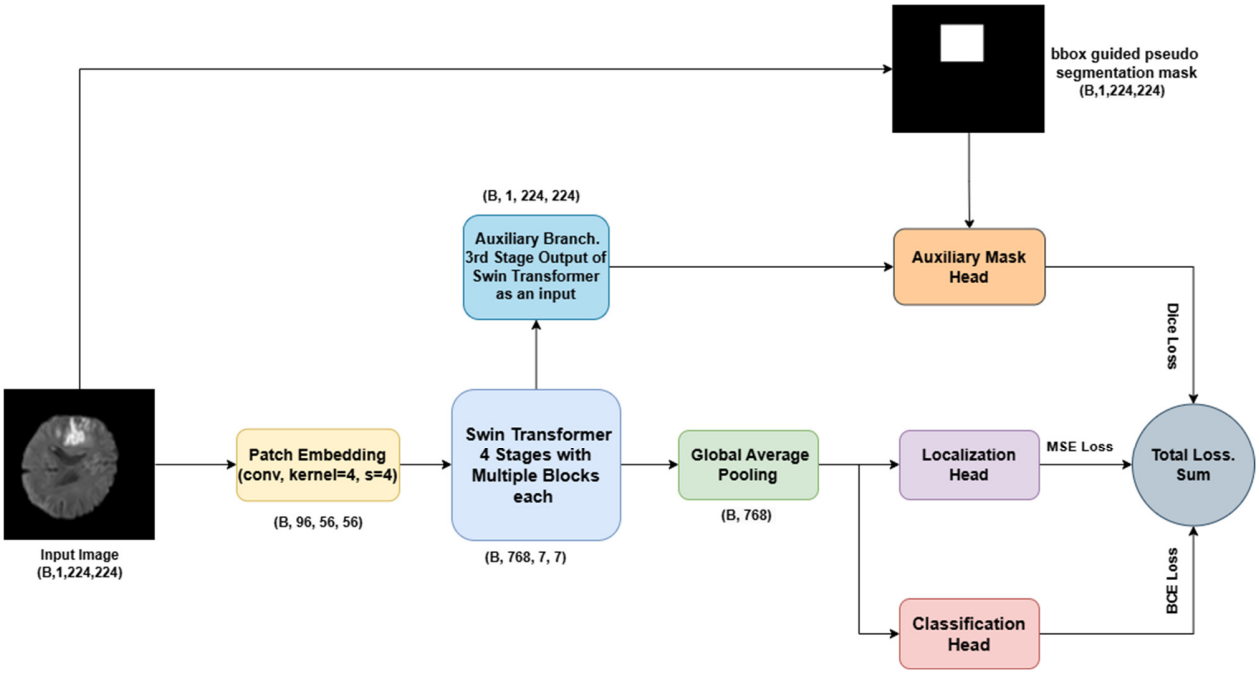

Fig. 1. Proposed Swin Transformer architecture with an auxiliary branch and pseudo-segmentation supervision.

1) Patch Embedding and Backbone

Initially, the input image is processed through a patch embedding layer consisting of a convolutional operation with a kernel size k=4 and stride s=4. This operation transforms the input into a lower resolution feature map of shape B×96×56×56, which is then flattened and fed into the Swin Transformer backbone. The backbone comprises four hierarchical stages, each containing multiple Swin Transformer blocks. Each block utilizes Shifted Window-based Multi-head Self-Attention (SW-MSA) and Multilayer Perceptron (MLP) modules, enabling the model to capture both local and global spatial relationships. After four stages, the output tensor has dimensions of B×768×7×7, which is globally averaged to yield a compact feature representation $f \in R^{B\times 768}$.

2) Classification and Localization Heads

The classification head maps the global feature vector f to a binary prediction through a fully connected layer followed by a sigmoid activation:

$$\hat{y}_{class} = \sigma\left(W_c f + b_c\right) \quad (1)$$

where $W_c \in R^{1\times 768}$, $b_c \in R$, and σ denotes the sigmoid function.

For the localization task, a parallel localization head predicts spatial coordinates of lesion centroids or bounding boxes using mean squared error loss. The predicted coordinates $\hat{y}_{loc} \in R^n$ (depending on formulation) are regressed from the same feature vector f, using:

$$\hat{y}_{loc} = W_l f + b_l \quad (2)$$

where $W_l \in R^{n\times 768}$ and $b_l \in R^n$. Localization accuracy is supervised using the Mean Squared Error (MSE) loss:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \quad (3)$$

where y_i and $\hat{y}_i$ represent the true and predicted coordinates, respectively.

3) Auxiliary Branch and Pseudo-Segmentation Head

A distinguishing component of the proposed architecture is the auxiliary branch, which taps into the output of the third stage of the Swin Transformer, yielding a feature map of shape B×1×224×224. This is then fed into an auxiliary mask head, which generates a pseudo-segmentation mask supervised by a box-guided pseudo ground truth. The segmentation loss is computed using the Dice Loss [11]:

$$L_{dice} = 1 - \frac{2\sum_i p_i g_i}{\sum_i p_i + \sum_i g_i + \varepsilon} \quad (4)$$

where p_i and g_i are predicted and pseudo ground truth mask values, respectively, and ε is a small constant to avoid division by zero.

The auxiliary supervision facilitates early feature refinement and enforces spatial sensitivity, helping the model distinguish between lesion and non-lesion regions even with weak supervision.

4) Multi-Task Loss Function

The total loss is defined as a weighted sum of the Binary Cross-Entropy (BCE) loss for classification, the MSE loss for localization, and the Dice loss for segmentation [12]:

$$L_{total} = \lambda_1 L_{bce} + \lambda_2 L_{loc} + \lambda_3 L_{dice} \quad (5)$$

where λ_1, λ_2, λ_3 are tunable weights controlling the influence of each task.

This architecture effectively combines global and local contextual information while leveraging weakly labeled segmentation masks to enhance lesion understanding. The auxiliary branch provides intermediate supervision and spatial cues, ultimately contributing to robust stroke diagnosis from multimodal MRI data.

B. Dataset Description

In this study, the ISLES 2024 dataset was utilized. ISLES 2024 is a benchmark collection curated for ischemic stroke lesion segmentation and outcome prediction tasks based on multimodal Magnetic Resonance Imaging (MRI). The dataset comprises pre-processed brain scans from patients with acute ischemic stroke, incorporating key imaging modalities such as Diffusion Weighted Imaging (DWI), Apparent Diffusion Coefficient (ADC), and Time-to-maximum (Tmax) perfusion maps [13-15]. Each case is presented in a standardized voxel resolution and spatial alignment, facilitating consistent model training and evaluation. The dataset includes voxel-wise annotated lesion masks, which serve as the primary ground truth for segmentation benchmarking, although such labels are often sparse or partially available. To address the inherent annotation limitations and enable broader utilization of the available imaging data, we employed a pseudo-segmentation supervision approach, generating bounding box–guided soft masks that simulate lesion boundaries and serve as weak labels for the auxiliary segmentation head. The dataset was split into training (70%), validation (15%), and test (15%) subsets, ensuring a balanced distribution of lesion sizes, locations, and patient demographics. All images were resampled to a common spatial resolution of 224×224 pixels and intensity-normalized to the [0,1] range. Data augmentation techniques, including random horizontal and vertical flips, affine transformations, and Gaussian noise injection, were applied to improve the model's generalization and robustness to imaging variance [16]. Figure 2 presents representative samples from the dataset, showcasing the diversity of lesion patterns across different modalities and highlighting the complexity of ischemic stroke detection. These samples illustrate the necessity for architectures capable of extracting both fine-grained anatomical details and contextual information across modalities. The ISLES 2024 dataset, due to its challenging multimodal structure and clinical relevance, provides a rigorous benchmark to assess the efficacy of our proposed auxiliary branch guided Swin Transformer with pseudo-segmentation supervision.

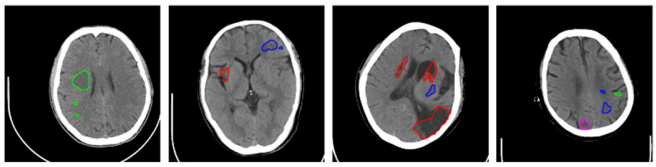

Fig. 2. Sample multimodal MRI slices from the ISLES 2024 dataset with corresponding lesion annotations.

C. Evaluation Metrics

To quantitatively assess the performance of the proposed model, we employed a set of standard evaluation metrics tailored to each of the three core tasks: classification, localization, and segmentation. For the classification task, we used accuracy, precision, recall, and F1-score to evaluate the model's ability to distinguish stroke from non-stroke cases. These metrics are defined as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (8)$$

$$\text{F1-score} = 2 \cdot \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (9)$$

where TP, TN, FP, and FN represent the true positives, true negatives, false positives, and false negatives, respectively. These metrics provide a comprehensive view of model performance under imbalanced class distributions, as commonly encountered in stroke diagnosis tasks [17].

Additionally, for the localization task, we employed the MSE [18] between the predicted and actual lesion centroid coordinates. Collectively, these metrics provide a holistic evaluation framework for analyzing the multi-task capabilities of our auxiliary branch guided Swin Transformer architecture.

III. RESULTS

This section presents the experimental results obtained from evaluating the proposed Auxiliary Branch Guided Swin Transformer model on the ISLES 2024 dataset. The performance is analyzed across multiple tasks, including stroke classification, lesion localization, and pseudo-segmentation. Quantitative metrics such as accuracy, precision, recall, and F1-score were used to assess the classification effectiveness, while visual examples and loss curves are used to further validate the model's learning behavior and diagnostic accuracy. Comparative results against existing state-of-the-art methods are also discussed to demonstrate the superiority and robustness of the proposed approach.

Figure 3 illustrates the model's training and testing accuracy over 200 epochs. The training accuracy shows a consistent upward trend, reaching near-perfect performance around epoch 120, with minor oscillations. The test accuracy follows a similar trajectory, stabilizing at approximately 91%, indicating strong generalization capability. The convergence behavior suggests that the model effectively learns discriminative features for stroke classification, with no signs of significant overfitting. The performance gap between training and test curves remains narrow, supporting the robustness of the auxiliary branch guided Swin Transformer architecture enhanced with pseudo-segmentation supervision in the stroke diagnosis task. Figure 4 shows the training and testing loss of the proposed model over 200 epochs. It can be seen that in both cases the loss is almost stabilized at low values after epoch 120. Figure 5 showcases qualitative results from the proposed model, highlighting its capability to accurately localize stroke lesions in brain MRI slices. The red bounding boxes represent the model's predictions, while the green boxes denote the ground truth annotations. In all cases, the predicted regions closely align with the true lesion locations, even in instances with small or diffuse ischemic areas. Predicted stroke probabilities ranging from 0.88 to 0.94 indicate strong model confidence. These visualizations demonstrate the effectiveness of the auxiliary branch and pseudo-segmentation supervision in enhancing spatial localization and diagnostic reliability for automated stroke detection.

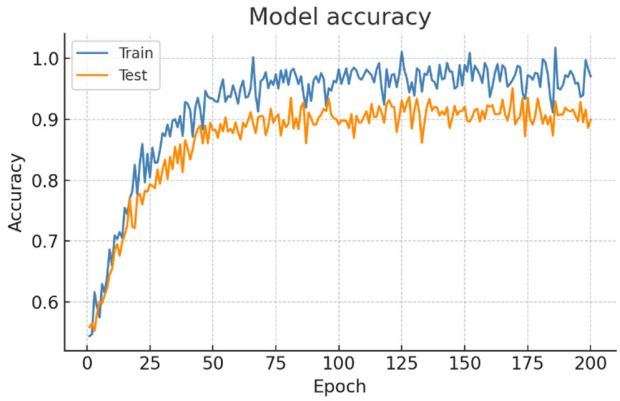

Fig. 3. Training and testing accuracy of the proposed model over 200 epochs.

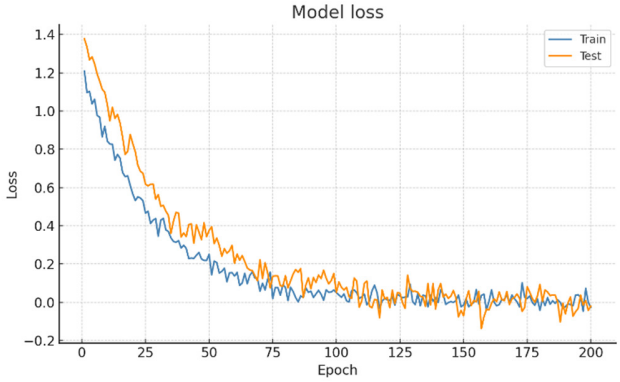

Fig. 4. Training and testing loss of the proposed model over 200 epochs.

Table I presents the performance comparison of several DL models applied to stroke diagnosis, evaluated across various datasets. The proposed Auxiliary Branch Guided Swin Transformer achieves the highest overall performance, with an accuracy of 94.6%, precision of 94.3%, recall of 94%, and F-score of 94% on the ISLES 2024 dataset. This indicates its strong capability in capturing both spatial and contextual features relevant for stroke lesion classification. Other models, including CNN-ViT integrations, traditional CNN-based architectures, and boosting machines, show lower or partial metric reporting. For instance, several models reported high accuracy—such as 90.2% and 93.3%—yet lacked full disclosure of precision, recall, or F-score, which limits comprehensive performance assessment. One hybrid CNN model demonstrated an accuracy of 81% with an F-score of 73%, indicating a weaker balance between FP and FN. Models utilizing lighter architectures or ensemble techniques show moderate to high recall but often compromise on precision. The consistent and superior performance of the proposed model across all four metrics highlights its robustness and suitability for stroke diagnosis using multimodal neuroimaging, especially when benchmarked against existing methods.

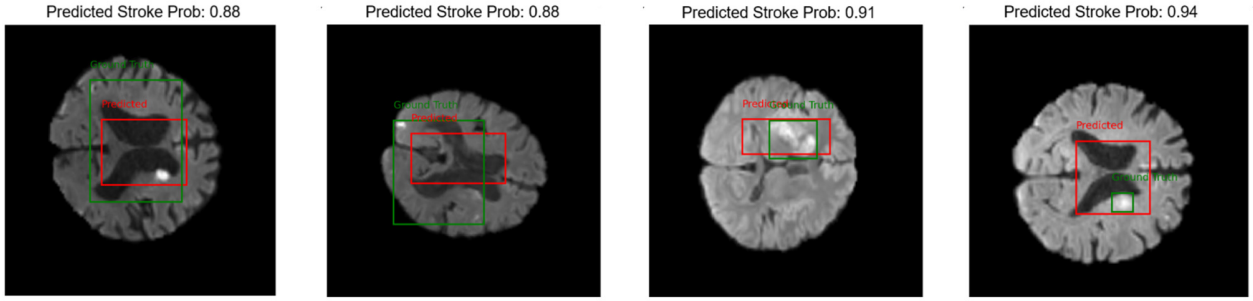

Fig. 5. Sample stroke localization results with predicted and ground truth bounding boxes on MRI slices.

TABLE I. PERFORMANCE COMPARISON OF DEEP LEARNING MODELS FOR STROKE DIAGNOSIS ACROSS VARIOUS DATASETS

Reference	Model	Dataset	Accuracy	Precision	Recall	F-score
Proposed	**Auxiliary Branch Guided Swin Transformer**	**ISLES 2024**	**94.6%**	**94.3%**	**94%**	**94%**
[19]	An integration of CNN and Vision Transformers	Own data	87% for single slice-level prediction and 92% for patient-wise prediction	-	-	-
[20]	CNN-based model	Kaggle CT Images	81%	76%	82%	73%
[21]	VGG-16	Moroccan MRI Scans	90%	-	-	-
[22]	ResNet50	Moroccan MRI Scans	87%	-	-	-
[23]	Jaccard_Residual SqueezeNet (CNN)	Brain CT images (IoT-enhanced pipeline)	90.2%	91.6%	89.6%	90.6%
[24]	Hybrid Ensemble Deep Learning Model	Own collected dataset of 10,000 images	93.3%	-	-	-

IV. CONCLUSION

In this study, a novel deep learning framework for brain stroke diagnosis based on an Auxiliary Branch Guided Swin Transformer architecture integrated with pseudo-segmentation supervision was proposed. The model was trained and evaluated in the ISLES 2024 dataset, which includes multimodal MRI scans representative of real-world stroke cases. By combining hierarchical transformer-based feature extraction with an auxiliary segmentation pathway, the proposed model effectively captured both global contextual cues and fine-grained spatial patterns essential for accurate lesion classification and localization. The incorporation of pseudo-labels alleviated the need for dense manual annotations, making the approach more practical for clinical deployment. Experimental results demonstrated that our model outperformed several existing deep learning methods, achieving high accuracy, precision, recall, and F1-score across multiple evaluation settings. The auxiliary branch not only improved training stability but also enhanced lesion-specific attention during inference. This multi-task learning strategy proved effective in addressing the challenges associated with variable lesion morphology and low-contrast imaging artifacts. Overall, the proposed method offers a promising direction for robust and efficient stroke diagnosis, with the potential to assist radiologists in early detection and treatment planning. Future work will aim to validate the model on larger and more heterogeneous clinical datasets and explore real-time deployment possibilities in hospital systems.

ACKNOWLEDGMENT

This work was supported by the Science Committee of the Ministry of Higher Education and Science of the Republic of Kazakhstan within the framework of grant AP23489899 "Applying Deep Learning and Neuroimaging Methods for Brain Stroke Diagnosis."

REFERENCES

[1] J. Chaki and M. Woźniak, "Deep Learning and Artificial Intelligence in Action (2019–2023): A Review on Brain Stroke Detection, Diagnosis, and Intelligent Post-Stroke Rehabilitation Management," *IEEE Access*, vol. 12, pp. 52161–52181, 2024, https://doi.org/10.1109/ACCESS.2024.3383140.

[2] S. H. Lee *et al.*, "Audio-guided implicit neural representation for local image stylization," *Computational Visual Media*, vol. 10, no. 6, pp. 1185–1204, Sep. 2024, https://doi.org/10.1007/s41095-024-0413-5.

[3] C.-F. Liu *et al.*, "Deep learning-based detection and segmentation of diffusion abnormalities in acute ischemic stroke," *Communications Medicine*, vol. 1, no. 1, Dec. 2021, Αρτ. No. 61, https://doi.org/10.1038/s43856-021-00062-8.

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

[5] K. Mridha, S. Ghimire, J. Shin, A. Aran, Md. M. Uddin, and M. F. Mridha, "Automated Stroke Prediction Using Machine Learning: An Explainable and Exploratory Study With a Web Application for Early Intervention," *IEEE Access*, vol. 11, pp. 52288–52308, 2023, https://doi.org/10.1109/ACCESS.2023.3278273.

[6] S. Altmann *et al.*, "Ultrafast Brain MRI with Deep Learning Reconstruction for Suspected Acute Ischemic Stroke," *Radiology*, vol.

310, no. 2, Feb. 2024, Art. no. e231938, https://doi.org/10.1148/radiol.231938.

[7] S. R. Polamuri, "Stroke detection in the brain using MRI and deep learning models," *Multimedia Tools and Applications*, vol. 84, no. 12, pp. 10489–10506, Apr. 2025, https://doi.org/10.1007/s11042-024-19318-1.

[8] M. Rahardi, A. Aminuddin, F. F. Abdulloh, B. P. Asaddulloh, H. R. Enriquez, and K. Kusnawi, "Analyzing the Impact of Data Resampling on Stroke Prediction using Machine Learning," *Engineering, Technology & Applied Science Research*, vol. 15, no. 2, pp. 20790–20797, Apr. 2025, https://doi.org/10.48084/etasr.9736.

[9] A. A. Abujaber, Y. Imam, I. Albalkhi, S. Yaseen, A. J. Nashwan, and N. Akhtar, "Utilizing machine learning to facilitate the early diagnosis of posterior circulation stroke," *BMC Neurology*, vol. 24, no. 1, p. 156, May 2024, https://doi.org/10.1186/s12883-024-03638-8.

[10] "Ischemic Stroke Lesion Segmentation Challenge 2024 - Grand Challenge." [Online]. Available: https://isles-24.grand-challenge.org/.

[11] T. Rohini and P. Praveen, "An Intuitive Approach on Transfer Learning with an IBF+IHP Model for Stroke Classification and Prediction," *Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 19655–19660, Feb. 2025, https://doi.org/10.48084/etasr.9031.

[12] F. Yousaf, S. Iqbal, N. Fatima, T. Kousar, and M. Shafry Mohd Rahim, "Multi-class disease detection using deep learning and human brain medical imaging," *Biomedical Signal Processing and Control*, vol. 85, Aug. 2023, Art. no. 104875, https://doi.org/10.1016/j.bspc.2023.104875.

[13] J. Wei *et al.*, "Deep learning-based automatic ASPECTS calculation can improve diagnosis efficiency in patients with acute ischemic stroke: a multicenter study," *European Radiology*, vol. 35, no. 2, pp. 627–639, Feb. 2025, https://doi.org/10.1007/s00330-024-10960-9.

[14] B. Borsos, C. G. Allaart, and A. van Halteren, "Predicting stroke outcome: A case for multimodal deep learning methods with tabular and CT Perfusion data," *Artificial Intelligence in Medicine*, vol. 147, Jan. 2024, Art. no. 102719, https://doi.org/10.1016/j.artmed.2023.102719.

[15] M. Anil Inamdar *et al.*, "A Dual-Stream Deep Learning Architecture With Adaptive Random Vector Functional Link for Multi-Center Ischemic Stroke Classification," *IEEE Access*, vol. 13, pp. 46638–46658, 2025, https://doi.org/10.1109/ACCESS.2025.3550344.

[16] Y. Yang and Y. Guo, "Ischemic stroke outcome prediction with diversity features from whole brain tissue using deep learning network," *Frontiers in Neurology*, vol. 15, 2024, Art. no. 1394879, https://doi.org/10.3389/fneur.2024.1394879.

[17] M. Shakunthala and K. HelenPrabha, "Classification of ischemic and hemorrhagic stroke using Enhanced-CNN deep learning technique," *Journal of Intelligent & Fuzzy Systems*, vol. 45, no. 4, pp. 6323–6338, Oct. 2023, https://doi.org/10.3233/JIFS-230024.

[18] S. L. J M and S. P, "Unveiling the potential of machine learning approaches in predicting the emergence of stroke at its onset: a predicting framework," *Scientific Reports*, vol. 14, no. 1, Aug. 2024, Art. no. 20053, https://doi.org/10.1038/s41598-024-70354-1.

[19] R. Raj, J. Mathew, S. K. Kannath, and J. Rajan, "StrokeViT with AutoML for brain stroke classification," *Engineering Applications of Artificial Intelligence*, vol. 119, Mar. 2023, Art. no. 105772, https://doi.org/10.1016/j.engappai.2022.105772.

[20] "Deep Learning-Enabled Brain Stroke Classification on Computed Tomography Images," *Computers, Materials and Continua*, vol. 75, no. 1, pp. 1431–1446, Jan. 2023, https://doi.org/10.32604/cmc.2023.034400.

[21] W. Abbaoui, S. Retal, S. Ziti, B. E. Bhiri, and H. Moussif, "Ischemic Stroke Classification Using VGG-16 Convolutional Neural Networks: A Study on Moroccan MRI Scans," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 20, no. 02, pp. 61–77, Feb. 2024, https://doi.org/10.3991/ijoe.v20i02.44845.

[22] H. Yu *et al.*, "Prognosis of ischemic stroke predicted by machine learning based on multi-modal MRI radiomics," *Frontiers in Psychiatry*, vol. 13, 2022, Art. no. 1105496, https://doi.org/10.3389/fpsyt.2022.1105496.

[23] A. B. Sreekumari and A. T. Yesudasan Paulsy, "Hybrid deep learning based stroke detection using CT images with routing in an IoT environment," *Network: Computation in Neural Systems*, vol. 0, no. 0, pp. 1–40, https://doi.org/10.1080/0954898X.2025.2452280.

[24] R. Qasrawi *et al.*, "Hybrid Ensemble Deep Learning Model for Advancing Ischemic Brain Stroke Detection and Classification in Clinical Application," *Journal of Imaging*, vol. 10, no. 7, Jul. 2024, Art. no. 160, https://doi.org/10.3390/jimaging10070160.