









*Original Research*

# The Role of Artificial Intelligence in Diagnosing Acute Ischemic Stroke Using Diffusion MRI: A Multicenter External Validation Study

Beyza Nur Kuzan<sup>1</sup>, Ali Abbasian Ardakani<sup>2</sup>, Mahmut Esat Aykan<sup>3</sup>, Mustafa Demir<sup>4</sup>, Servan Yaşar<sup>3</sup>, Mehmet Semih Çakır<sup>5</sup>, Afshin Mohammadi<sup>6</sup>, Taha Yusuf Kuzan<sup>3,\*</sup>

<sup>1</sup>Department of Radiology, Kartal Dr. Lütfi Kırdar City Hospital, 34865 Istanbul, Türkiye

<sup>2</sup>Department of Radiology Technology, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, 1971653313 Tehran, Iran

<sup>3</sup>Department of Radiology, Sancaktepe Şehit Prof. Dr. İlhan Varank Training and Research Hospital, University of Health Sciences, 34785 Istanbul, Türkiye

<sup>4</sup>Department of Radiology, Ümraniye Training and Research Hospital, 34764 Istanbul, Türkiye

<sup>5</sup>Department of Radiology, Faculty of Medicine, Istanbul University, 34093 Istanbul, Türkiye

<sup>6</sup>Department of Radiology, Faculty of Medicine, Urmia University of Medical Science, 5714783734 Urmia, Iran

\*Correspondence: [tykuzan@gmail.com](mailto:tykuzan@gmail.com) (Taha Yusuf Kuzan)

Academic Editors: Parisa Gazerani and Bettina Platt

Submitted: 4 December 2025 Revised: 26 February 2026 Accepted: 4 March 2026 Published: 21 April 2026

## Abstract

**Background:** Acute ischemic stroke is a time-sensitive medical emergency requiring rapid and accurate diagnosis to improve patient outcomes. While Diffusion-Weighted Imaging (DWI) on magnetic resonance imaging (MRI) is highly sensitive, artificial intelligence (AI) offers a potential solution to enhance diagnostic speed and accuracy. In this study we aimed to evaluate and externally validate a DWI-MRI-based deep learning model for automated stroke detection and compare its multicenter diagnostic performance with expert radiologists to determine generalizability and clinical utility. **Methods:** This retrospective study involved 732 patient cases (acute ischemic stroke and non-stroke controls) from three different centers. A deep convolutional neural network (CNN) model was developed, trained, and internally validated using data from center 1 (n = 452 for training, n = 80 for validation). The model's generalizability was then tested using independent external validation datasets from center 2 (n = 100) and center 3 (n = 100). The model's diagnostic performance (sensitivity, specificity, accuracy, and area under the receiver operating characteristic (ROC) curve (AUC)) was systematically compared with that of three expert radiologists. **Results:** The deep learning model demonstrated excellent diagnostic performance. In the internal validation, the model achieved 100% sensitivity, 100% specificity, and 100% accuracy (AUC = 1.000). Crucially, it maintained high performance on the external validation cohorts, achieving 100% sensitivity, 98% specificity, and 99% accuracy for both center 2 (AUC = 0.987) and center 3 (AUC = 0.986). This performance was comparable with the expert radiologists, who also achieved high accuracy across all datasets. Visualization techniques (gradient-weighted class activation map (Grad-CAM)) confirmed that the AI model focused on the correct pathological regions when making its classifications. **Conclusions:** The DWI-MRI-based deep learning model provides high and reliable diagnostic accuracy for acute ischemic stroke, with performance comparable with that of expert radiologists. Its robust performance across multicenter data highlights its potential as a dependable decision-support tool in emergency departments, especially in settings with limited specialist availability, to facilitate faster and more consistent stroke diagnoses.

**Keywords:** brain ischemia; artificial intelligence; deep learning; diffusion magnetic resonance imaging; diagnosis; computer-assisted

## 1. Introduction

The advent of artificial intelligence (AI) has profoundly transformed numerous disciplines, including the critical field of healthcare. AI has rapidly emerged as a transformative and beneficial tool in the crucial diagnosis and comprehensive management of acute stroke. Stroke is a health problem that is of significant concern. Stroke is divided into two primary types: ischemic and hemorrhagic. It is the fifth leading cause of death in the United States and a major cause of disability worldwide [1]. Each year, approximately 800,000 individuals in the USA face the devastating reality of a new or recurrent stroke, underscoring the urgent need for rapid, accurate, and effective diagnostic procedures to significantly improve clinical outcomes for

patients [2,3]. The concept of the “Golden Hour” is a term used to emphasize the critical and time-sensitive nature of timely intervention. It is during this limited time period that various treatment options can most effectively mitigate and alleviate the enduring consequences of a stroke [4,5]. Nevertheless, the recognition of acute stroke symptoms in emergency medical settings presents a considerable and complex challenge. This is due to the intricacies and time-sensitive demands of emergency care. Consequently, delays in treatment can result in the loss of opportunities for effective and timely interventions that could ultimately save lives and improve the quality of recovery for stroke patients [1,6].

Multimodal imaging approaches—including non-contrast Computed Tomography (CT), CT angiography, CT perfusion, and the diffusion-weighted imaging (DWI)—



fluid-attenuated inversion recovery (FLAIR) mismatch—are central to contemporary acute ischemic stroke evaluation and management. Among these modalities, DWI has a distinctive role, as it represents the most sensitive magnetic resonance imaging (MRI) sequence for the detection of acute ischemic infarction once MRI is performed. Consequently, DWI is routinely incorporated into the diagnostic workflow of acute ischemic stroke and serves as a critical imaging modality in clinical practice [7]. Inadequate or delayed diagnosis can lead to poorer clinical outcomes, emphasizing the necessity for robust clinical support systems in emergency departments [8]. The ongoing Fourth Industrial Revolution has furthered the application of AI technologies, particularly in the field of medical imaging [9]. The incorporation of AI into stroke diagnosis has been demonstrated to enhance diagnostic efficiency and hold the potential to improve patient survival rates and functional recovery outcomes [10–12]. Therefore, the deployment of AI in acute stroke diagnosis is a promising frontier in the pursuit of more effective and timely patient care in emergency departments with limited expertise in stroke management [13]. While emergency MRI availability is not universal, many comprehensive and tertiary stroke centers routinely incorporate MRI-DWI into their acute stroke protocols, especially when diagnostic uncertainty exists or when stroke mimics must be excluded. In this context, an AI-based system trained exclusively on DWI may provide robust lesion detection and serve as a valuable decision-support tool in MRI-equipped emergency departments rather than as a replacement for multimodal imaging strategies.

The progressive development of AI in the field of stroke diagnosis signifies a pivotal advancement in medical technology, with the potential to enhance diagnostic efficiency and contribute to the preservation of lives through early intervention strategies. In order to achieve a desirable level of generalizability for any AI system, it is necessary to validate it using external databases with different clinical settings, imaging protocols, and populations. Additionally, a comparative analysis with radiologists is necessary to ascertain the relative efficacy of these systems in actual clinical settings [12,14]. In the present study, the validity of a DWI-MRI-based deep learning model used for acute stroke detection is evaluated. In addition, the performance of the model was assessed through the analysis of databases from multiple centers, with the goal of determining its efficacy in routine clinical practice.

## 2. Patient Selection and Methods

This retrospective study investigates the results of the application of an AI system based on diffusion MRI findings in acute ischemic stroke and non-stroke patients. The model was developed and validated both internally and externally.

### 2.1 Patient Selection

The study was approved by the institutional review board (Approval number: 2023-91) and the requirement for informed consent was waived. This study's participants comprised patients with acute ischemic stroke and control patients evaluated in the emergency department with suspected acute ischemic stroke from three different centers in Istanbul, Türkiye. The centers included Sancaktepe Şehit Prof. Dr. İlhan Varank Training and Research Hospital (center 1), Ümraniye Training and Research Hospital (center 2), and Kartal Dr. Lütfi Kırdar City Hospital (center 3).

Patients presenting to the emergency department with acute neurological symptoms suspicious of ischemic stroke were screened for inclusion. All patients underwent non-contrast CT and diffusion-weighted MRI as part of routine clinical evaluation. Cases labeled as acute ischemic stroke were defined by the presence of diffusion restriction on DWI consistent with acute infarction and confirmed through retrospective consensus evaluation by expert radiologists together with the final clinical diagnosis documented in the medical records. The control group consisted of patients who presented with suspected acute stroke but in whom acute ischemic stroke was excluded based on imaging findings (absence of diffusion restriction on DWI and no acute ischemic findings on CT), clinical evaluation, and the final diagnosis recorded in the hospital records. To improve dataset reliability and reduce potential confounding factors, patients were excluded if they had significant imaging artifacts, prior stroke, previous neurosurgery, intracranial mass lesions, or pre-existing neurological diseases such as multiple sclerosis or infection. In addition, lacunar infarctions were excluded to focus on clearly visible acute ischemic lesions suitable for model training.

A total of 732 patients were included from three centers. From center 1, 532 cases were included and randomly split at the patient level into training ( $n = 452$ ) and internal validation ( $n = 80$ ) cohorts. Data from center 2 ( $n = 100$ ) and center 3 ( $n = 100$ ) were used exclusively for external validation. The datasets were balanced with respect to stroke and control cases.

### 2.2 Data Acquisition and Preprocessing

MRI was performed in center 1 and 2 with a 1.5-Tesla MRI machine (Signa Explorer; General Electric Healthcare, Milwaukee, WI, USA) and in center 3 with a 1.5-Tesla MRI machine (Ingenia MR 5300, Philips Healthcare, Best, Netherlands). The diffusion MRI parameters on the Signa Explorer were as follows: in-plane resolution, 1.5 mm × 1.5 mm; slice thickness, 5 mm; echo time (TE), 86.6 ms; repetition time (TR), 5000 ms; partial Fourier 6/8. Two b-shells were acquired with b-values of 1000 s/mm<sup>2</sup> and 2000 s/mm<sup>2</sup>. The diffusion MRI parameters on the Philips MRI machine were as follows: the parameters are TR/TE 5598 ms/105 ms; flip angle 90°; slice thickness 5 mm; Field-of-view (FOV) 24 × 24 cm; matrix size 256 × 256.

For model input, the network was trained using the trace diffusion-weighted images generated by the scanner from the clinical DWI sequence, which correspond to the images routinely interpreted in clinical practice. Apparent diffusion coefficient (ADC) maps and individual b-value images were not used as separate inputs for the model.

The final reference standard for AI training and testing was established through a retrospective consensus review by two expert radiologists who independently evaluated all patient cases while blinded to the AI model outputs. In cases of disagreement, a third senior radiologist provided adjudication, and the consensus decision was used as the final label. Imaging findings were interpreted together with the clinical diagnosis documented in the medical records. This expert-based annotation process ensured that model labels were independent of the initial enrollment diagnosis and served as a reliable reference standard for performance evaluation.

A two-step preprocessing approach was applied to reduce discrepancies in image appearance resulting from the multicenter nature of the study. This step ensures that inter-scan and intra-scan variability are accounted for and improves the reproducibility of the proposed AI model. All images from different centers were standardized using image histogram matching. The reference histogram was computed from images in the training dataset. For each training image, we first applied simple intensity clipping (0.5th–99.5th percentile) to reduce the influence of extreme outliers, and then calculated a 256-bin intensity histogram. The final reference histogram was obtained by averaging these histograms bin-wise. This reference was subsequently used for histogram matching of all images in the training, internal validation, and external validation sets. Then, all MRI slices were normalized using Z-score normalization, which standardizes the intensity distribution across subjects and provides a consistent input scale for the deep learning model.

### 2.3 Artificial Intelligence Study

The pre-processed diffusion MRI images were used as input to the deep learning-based model for differentiation of acute ischemic stroke and normal cases. In the diagnostic process, image features are automatically extracted using a deep learning algorithm, and the most relevant features are identified for the classification task. The model was trained using slice-level DWI images. However, all dataset splitting was performed at the patient level to ensure independence, and no slices from the same patient were included in more than one dataset. For performance evaluation, slice-level predictions were aggregated at the patient level using a majority-voting approach, whereby a patient was classified as acute ischemic stroke if the majority of slices were predicted as positive. The model simultaneously performs the classification process based on the best and informative deep features. In deep learning, image fea-

tures are extracted and analyzed hierarchically from raw data to high-level concepts, in different layers of the network. The architecture of the proposed convolutional neural network (CNN) is provided in Fig. 1. MRI images were resized to  $140 \times 140$  pixels and were fed into the model. The model consists of two convolutional and seven residual blocks. The convolutional block (ConvBlock) starts with a convolutional layer followed by a batch normalization (BatchN) and a rectified linear unit (ReLU) layer. The output of the first ConvBlock is split into two branches in the residual block (ResBlock). The first branch consists of a convolution layer followed by a BatchN layer, a ReLU activation, a convolutional layer, and a BatchN layer. Then, the outputs of this branch and the skip connection are added using an addition layer. On the other hand, the last two ResBlocks consist of a convolution layer followed by BatchN and ReLU layers with a skip connection. Between all ResBlocks, a pooling layer is put to reduce the dimension of data and preserve meaningful features. The size of filters in all convolution layers is  $3 \times 3$ . At the end, a global max (GlobMax) layer followed by two dense and dropout layers is used to prevent overfitting (Fig. 1).

In addition to dropout layer, an augmentation technique was used to prevent overfitting and increase generalizability. In this study, MRI images were randomly rotated ( $-30, 30$  degrees) and scaled (0.6, 1). The model was trained using stochastic gradient descent (sgdm) optimizer, cross-entropy loss function, initial learning rate of 0.001, validation frequency of 2, and L2 regularization of 0.001. Before each training epoch, the training dataset was shuffled to prevent learning order-specific patterns and improve convergence of the model. Hyperparameters were determined through preliminary empirical testing aimed at achieving stable training behavior and preventing overfitting. A formal hyperparameter optimization procedure (e.g., grid search or automated tuning) was not performed in this study.

The classification threshold was defined according to the training set using Receiver Operating Characteristic (ROC) analysis and subsequently fixed. This same threshold was applied without modification to the validation datasets when calculating sensitivity, specificity, and accuracy, ensuring consistent and unbiased performance evaluation.

The dataset from center 1 was split into 85% for training and tuning the model and 15% for the validation phase. Data from centers 2 and 3 were considered for the external validation step. In addition to deep learning, the internal and external validation datasets were reviewed by three expert radiologists in neuroimaging with 7 years (center 1), 5 years (center 2), and 7 years (center 3) to diagnose acute ischemic stroke cases. No cross-validation or permutation testing was performed in this study.

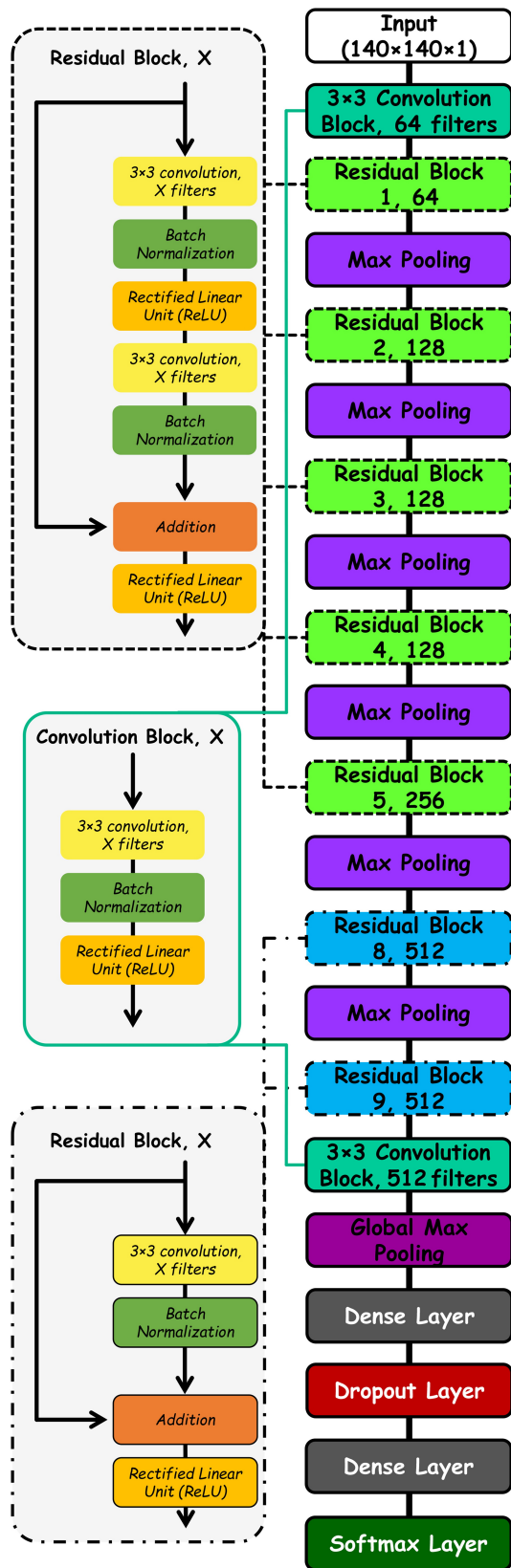


Fig. 1. The proposed deep learning model used for stroke detection.

## 2.4 Statistical Analysis and Performance Evaluation

To evaluate the proposed model in internal and external validation phases, the following metrics were determined: Sensitivity (Sen), Specificity (Sp), Accuracy (Acc), and area under the ROC curve (AUC). In this study, acute ischemic stroke and normal cases were considered as positive and negative, respectively. Thereby, normal cases misdiagnosed as stroke and stroke cases misdiagnosed as normal were classified as false positive and false negative, respectively. ROC curve analysis was used to determine the AUC with 95% confidence interval (CI) and evaluate the overall performance of the model on both internal and external validation datasets. All mentioned metrics were also determined for the radiologists in diagnosing acute ischemic stroke. In addition to these metrics, the gradient-weighted class activation map (Grad-CAM) technique was utilized to evaluate if the model considered and learned informative features. AUCs of the proposed deep learning model and three radiologists were compared using the DeLong test. A  $p$ -value less than 0.05 was considered significant. All statistical analyses were performed using SPSS software (version 24, SPSS Inc.; Chicago, IL, USA).

## 3. Results

### 3.1 Patients

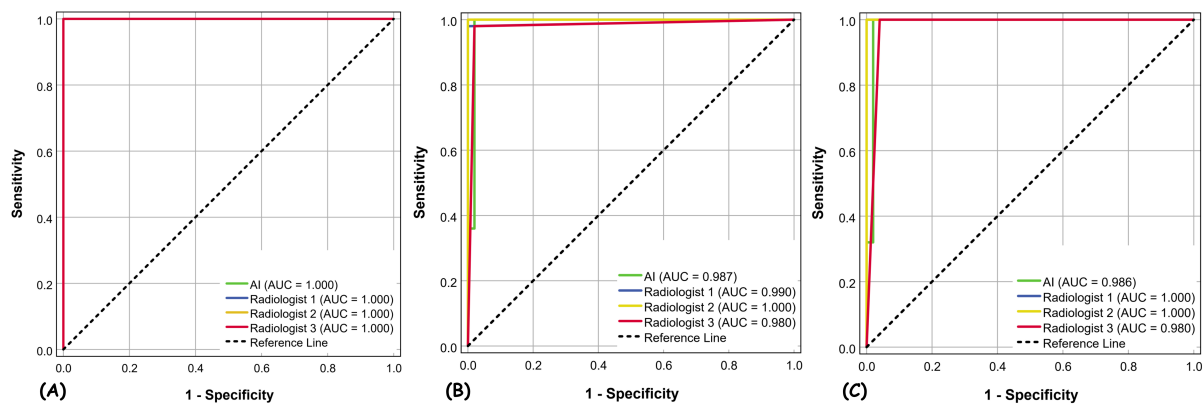
A total of 732 cases from three centers were included in this study. 452 cases (226 normal and 226 stroke) from center 1 and the remaining cases (40 normal and 40 stroke) were used for training and tuning, and validation of the model, respectively. In addition, 200 cases from center 2 (50 normal, 50 stroke) and center 3 (50 normal, 50 stroke) were considered for external validation.

### 3.2 Deep Learning Model Performance on Multi-Center Cohorts

The proposed deep-learning model demonstrated excellent performance in diagnosing acute ischemic stroke patients with a sensitivity, specificity, and accuracy of 100% in the internal validation dataset. In addition, it could differentiate acute ischemic stroke patients from healthy controls with a sensitivity of 100%, specificity of 98% and accuracy of 99% in centers 2 and 3. ROC curves and AUC values with 95% CI and confusion matrices of the model are presented in Fig. 2 and Table 1, respectively. The Grad-CAMs of the proposed model for some cases in the three centers are shown in Fig. 3.

### 3.3 Radiologists' Performance

All three radiologists could diagnose acute ischemic stroke patients with sensitivity, specificity and accuracy of 100% in the internal validation dataset. On the other hand, only radiologist 2 diagnosed stroke with sensitivity, specificity, and accuracy of 100% in both centers 2 and 3. The overall performance (AUC) of radiologist 1 in centers 2 and 3 were 0.990 (98% sensitivity, 100% specificity) and 1



**Fig. 2. Diagnostic performance of the proposed artificial intelligence (AI) model and three radiologists for internal validation (A), external validation 1 (B) and external validation 2 (C).**

(100% sensitivity, 100% specificity), respectively. Finally, the AUC value of radiologist 3 in center 2 and 3 were 0.980 (98% sensitivity, 98% specificity) and 0.980 (100% sensitivity, 96% specificity), respectively. ROC curves and the detailed performance of radiologists with the corresponding confusion matrices are presented in Table 1 and Fig. 2, respectively.

### 3.4 Artificial Intelligence (AI) and Radiologists' Performance Comparison

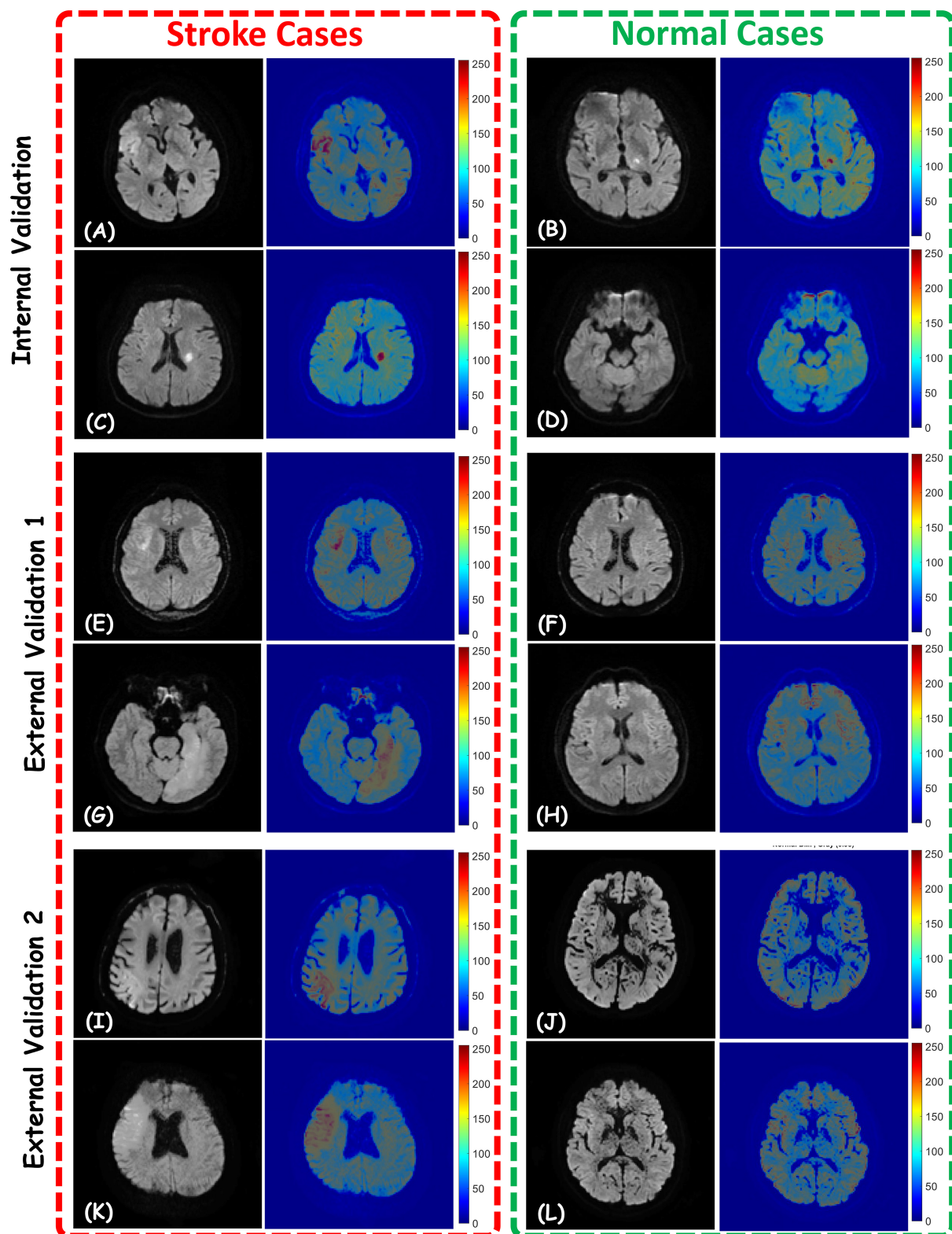
Fig. 4 shows some sample MRI images with the corresponding diagnosis of the proposed AI model and three radiologists. In addition, Grad-CAMs of each case are presented to show how the proposed model conceived MRI images and made differential diagnosis. In addition, the comparison results of the ROC curves using the DeLong test are summarized in Table 2. Notably, no significant differences were observed between the methods (the deep learning model and the three radiologists), underscoring the applicability of the proposed deep learning model.

## 4. Discussion

In this multicenter study, a DWI-MRI-based deep learning model demonstrated high diagnostic accuracy for acute ischemic stroke, as evidenced by the model's performance on both internal and external validation datasets. The model demonstrated a 100% accuracy rate in the internal validation and up to a 99% accuracy rate in external centers, exhibiting performance comparable to that of radiologists. The findings indicate that AI-based decision support systems can play a substantial role in facilitating rapid and reliable diagnosis, particularly within emergency departments where specialist neuroradiology support is scarce or in high-volume settings. Although perfect performance in internal validation may raise concerns regarding overfitting, the consistent results observed across two independent external validation cohorts support the robustness and generalizability of the proposed model.

Recent studies have demonstrated that AI has been shown to yield results similar to those of radiologists in stroke imaging (non-contrast CT, CT angiography, CT perfusion, MR angiography, DWI) for lesion detection, Alberta stroke program early CT score (ASPECTS) scoring, and detection of large vessel occlusions [15,16]. The development of deep learning models tailored for DWI has yielded high levels of sensitivity and specificity, achieved by minimizing error rates in the detection of small lesions. As stated by Liu *et al.* [16], the utilization of deep learning algorithms on DWI data yielded diagnostic accuracies that were comparable to those attained by radiologists with extensive experience. Similarly, Abedi *et al.* [1] underscored the potential of AI-supported systems to expedite emergency department workflow and reduce mortality. The present study lends support to these findings in the existing literature. The multicenter design of the study serves to strengthen the model's generalizability. The high performance demonstrated, even with data obtained from different devices and protocols, highlights the effectiveness of the preprocessing (histogram matching, normalization) and data augmentation methods used.

Accurate and prompt diagnosis of acute stroke is paramount for effective utilization of time windows for intravenous thrombolysis and mechanical thrombectomy. AI-supported automatic analyses have the potential to enhance safety measures, especially during the triage and diagnostic processes within emergency departments. The existing literature indicates that the implementation of automated software, such as RAPID, has been demonstrated to enhance clinical outcomes [17–19]. From a practical perspective, several challenges must be addressed before clinical deployment, including seamless integration with PACS, regulatory approval, and ensuring real-time or near real-time processing in emergency settings. Additionally, robust validation across diverse imaging protocols and scanners is essential to support safe and reliable implementation.



**Fig. 3. Visualization of Model Predictions using Gradient-weighted Class Activation Maps (Grad-CAMs).** Representative diffusion-weighted imaging (DWI) slices and corresponding Grad-CAMs for stroke (left columns) and normal (right columns) cases across three distinct datasets: Internal Validation (A–D), External Validation 1 (E–H), and External Validation 2 (I–L). All cases were correctly diagnosed by the proposed AI model.

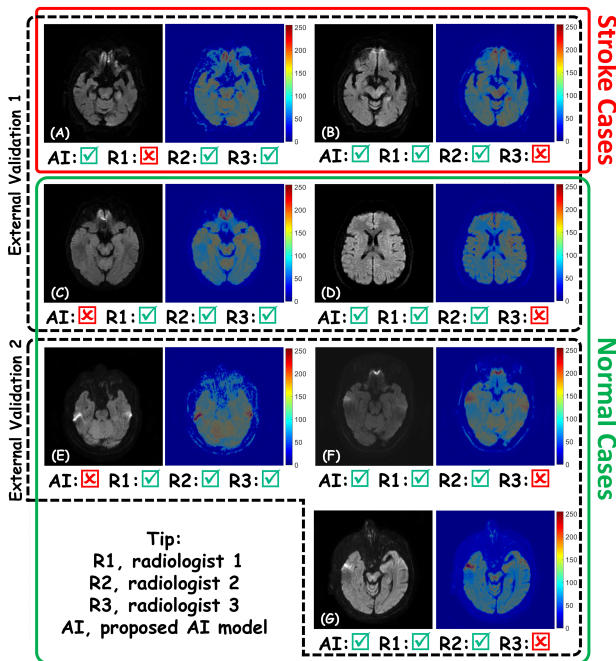
**Table 1. Diagnostic performance of the proposed artificial intelligence (AI) model and three radiologists based on each center.**

Dataset	Mode	Sen (95% CI)	Spc (95% CI)	Acc (%)	AUC (95% CI)	Confusion matrix		Predicted label
						True label		
						Stroke	Control	
Internal validation (center 1)	Deep Model	100 (91.2–100.0)	100 (91.2–100.0)	100.00	1.000	40	0	Stroke
						0	40	Control
	Radiologist 1	100 (91.2–100.0)	100 (91.2–100.0)	100.00	1.000	40	0	Stroke
						0	40	Control
	Radiologist 2	100 (91.2–100.0)	100 (91.2–100.0)	100.00	1.000	40	0	Stroke
						0	40	Control
Radiologist 3	100 (91.2–100.0)	100 (91.2–100.0)	100.00	1.000	40	0	Stroke	
					0	40	Control	
External validation 1 (center 2)	Deep Model	100 (92.9–100.0)	98.00 (89.1–99.9)	99.00	0.987 (0.962–1.000)	50	1	Stroke
						0	49	Control
	Radiologist 1	98.00 (89.1–99.9)	100 (92.9–100.0)	99.00	0.990 (0.967–1.000)	49	0	Stroke
						1	50	Control
	Radiologist 2	100 (92.9–100.0)	100 (92.9–100.0)	100.00	1.000	50	0	Stroke
						0	50	Control
Radiologist 3	98.00 (89.1–99.9)	98.00 (89.1–99.9)	98.00	0.980 (0.948–1.000)	49	1	Stroke	
					1	49	Control	
External validation 2 (center 3)	Deep Model	100 (92.9–100.0)	98.00 (89.1–99.9)	99.00	0.986 (0.960–1.000)	50	1	Stroke
						0	49	Control
	Radiologist 1	100 (92.9–100.0)	100 (92.9–100.0)	100.00	1.000	50	0	Stroke
						0	50	Control
	Radiologist 2	100 (92.9–100.0)	100 (92.9–100.0)	100.00	1.000	50	0	Stroke
						0	50	Control
Radiologist 3	100 (92.9–100.0)	96.00 (87.3–99.4)	98.00	0.980 (0.948–1.000)	50	2	Stroke	
					0	48	Control	

Sen, sensitivity; Spc, specificity; Acc, accuracy; AUC, area under the ROC curve; CI, confidence interval.

**Table 2. Comparison of diagnostic performance of the proposed artificial intelligence (AI) model and three radiologists based on each center.**

Center	Method	Deep model	Radiologist 1	Radiologist 2	Radiologist 3
Internal validation (center 1)	Deep Model	-	1.000	1.000	1.000
	Radiologist 1	1.000	-	1.000	1.000
	Radiologist 2	1.000	1.000	-	1.000
	Radiologist 3	1.000	1.000	1.000	-
External validation 1 (center 2)	Deep Model	-	0.860	0.319	0.704
	Radiologist 1	0.860	-	0.312	0.566
	Radiologist 2	0.319	0.312	-	0.153
	Radiologist 3	0.704	0.566	0.153	-
External validation 2 (center 3)	Deep Model	-	0.311	0.324	0.735
	Radiologist 1	0.311	-	1.000	0.147
	Radiologist 2	0.324	1.000	-	0.147
	Radiologist 3	0.735	0.147	0.147	-



**Fig. 4. Analysis of diagnostic discrepancies.** DWI slices and Grad-CAMs of cases with discordant findings between the AI model and three radiologists (R1, R2, R3). (A,B) Stroke cases: The AI model correctly identifies subtle lesions missed by R1 or R3. (C–G) Normal cases: (C,E) represent AI false positives where the model was misled by artifacts; (D,F,G) represent radiologist false positives where the AI correctly identified the scans as normal.

It should be emphasized that the proposed model is not intended to replace multimodal stroke imaging workflows but rather to complement existing diagnostic pathways in MRI-equipped centers. The reliance on DWI alone may limit applicability in emergency settings where MRI access is restricted; however, in centers with rapid MRI capabil-

ity, automated DWI-based stroke detection may enhance diagnostic confidence, reduce interpretation time, and support clinical decision-making, particularly in high-volume or resource-limited environments.

The Gradient-weighted Class Activation Mapping (Grad-CAM) method, as implemented in our study, demonstrated that the model focused on lesion areas when making decisions. Grad-CAM is a visualization technique that generates class-specific activation maps using the gradient information in the last convolutional layer to explain the decision-making processes of deep convolutional neural networks (CNNs). In CNN models trained for ischemic stroke detection in brain MRI images, the Grad-CAM method enables insightful analysis by localizing the spatial regions that guide the model’s prediction and offers the opportunity to assess the clinical relevance of the decision mechanism. Specifically, when contrasting ischemic lesions and normal tissues in diffusion-weighted images, heat maps derived from Grad-CAM can directly illustrate the location and extent of the lesion. In a study by Lai *et al.* [20], Grad-CAM outputs, which allow for visual assessment of the location and extent of lesions in diffusion-weighted images, are used as an important tool to evaluate whether the model correctly processes pathological regions. Other recent studies in the literature have also shown that Grad-CAM improves both model performance and explainability in automatic ischemic stroke detection systems [21,22]. Consequently, Grad-CAM not only fosters classification accuracy but also assumes a pivotal role in the implementation of AI-based clinical decision support systems. In a small number of cases, Grad-CAM visualizations highlighted regions adjacent to, rather than directly overlapping with, radiologists’ annotated lesions. Such discrepancies may reflect sensitivity to subtle diffusion changes, partial volume effects, or image noise, and underscore the complementary role of AI-based attention maps rather than direct lesion delineation.

The retrospective design of this study may introduce selection bias. Another major limitation is the exclusion of patients with prior stroke and lacunar infarctions, which may limit real-world applicability. Chronic diffusion abnormalities following previous stroke can overlap with acute lesions, complicating reliable labeling, while the slice-level classification approach used in this study is less sensitive to very small lacunar infarcts. In addition, stroke subtypes and the topographic distribution of ischemic lesions (e.g., vascular territories or lobar location) were not systematically analyzed, as etiological and anatomical localization were beyond the scope of this detection-focused study. Accordingly, the reported diagnostic performance should be interpreted as the model's ability to detect DWI-visible diffusion abnormalities rather than as a fully comprehensive clinical diagnosis of acute ischemic stroke. The exclusive use of DWI without integration of other imaging modalities such as CTA, CTP, or DWI-FLAIR mismatch may restrict generalizability to centers without emergency MRI availability. Future studies should incorporate multimodal imaging data, lesion-level segmentation, and prospective designs to evaluate workflow impact and clinical outcomes, as well as broader international external validation to enhance robustness and clinical applicability.

## 5. Conclusions

In summary, the present study demonstrated that the DWI MRI-based deep learning model provides diagnostic accuracy comparable to that of radiologists in the diagnosis of acute ischemic stroke and can be reliably applied in multicenter settings. In the future, the integration of AI systems into the clinical workflow for acute stroke care may be facilitated by prospective and multimodal studies.

## Availability of Data and Materials

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

## Author Contributions

TYK and AAA designed the study. BNK, MEA, SY, MD collected the data. AAA, MSC, AM, TYK analyzed the data and prepared the figures. BNK, MEA, SY, MD and MSC drafted the manuscript. TYK, AAA, AM brought major revisions in significant proportions of the manuscript. TYK and AM supervised the study, provided important intellectual input, and critically revised the manuscript for important intellectual content. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

## Ethics Approval and Consent to Participate

The study was conducted in accordance with the Declaration of Helsinki. Ethical approval was granted by the Institutional Review Board of the University of Health Sciences, Sancaktepe Şehit Prof. Dr. İlhan Varank Training and Research Hospital, Istanbul, Türkiye (Approval No: 2023-91). Due to the retrospective nature of the study, the requirement for informed consent was waived.

## Acknowledgment

Not applicable.

## Funding

This research received no external funding.

## Conflict of Interest

The authors declare no conflict of interest.

## References

- [1] Abedi V, Khan A, Chaudhary D, Misra D, Avula V, Mathrawala D, *et al.* Using artificial intelligence for improving stroke diagnosis in emergency departments: a practical framework. *Therapeutic Advances in Neurological Disorders.* 2020; 13: 1756286420938962. <https://doi.org/10.1177/1756286420938962>.
- [2] Roger VL, Go AS, Lloyd-Jones DM, Adams RJ, Berry JD, Brown TM, *et al.* Heart disease and stroke statistics—2011 update: a report from the American Heart Association. *Circulation.* 2011; 123: e18–e209. <https://doi.org/10.1161/CIR.0b013e3182009701>.
- [3] Latchaw RE, Alberts MJ, Lev MH, Connors JJ, Harbaugh RE, Higashida RT, *et al.* Recommendations for imaging of acute ischemic stroke: a scientific statement from the American Heart Association. *Stroke.* 2009; 40: 3646–3678. <https://doi.org/10.1161/STROKEAHA.108.192616>.
- [4] Saver JL, Smith EE, Fonarow GC, Reeves MJ, Zhao X, Olson DM, *et al.* The “golden hour” and acute brain ischemia: presenting features and lytic therapy in >30,000 patients arriving within 60 minutes of stroke onset. *Stroke.* 2010; 41: 1431–1439. <https://doi.org/10.1161/STROKEAHA.110.583815>.
- [5] Advani R, Naess H, Kurz MW. The golden hour of acute ischemic stroke. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine.* 2017; 25: 54. <https://doi.org/10.1186/s13049-017-0398-5>.
- [6] Zhao K, Zhao Q, Zhou P, Liu B, Zhang Q, Yang M. Can Artificial Intelligence Be Applied to Diagnose Intracerebral Hemorrhage under the Background of the Fourth Industrial Revolution? A Novel Systemic Review and Meta-Analysis. *International Journal of Clinical Practice.* 2022; 2022: 9430097. <https://doi.org/10.1155/2022/9430097>.
- [7] Wintermark M, Sanelli PC, Albers GW, Bello J, Derdeyn C, Hetsch SW, *et al.* Imaging recommendations for acute stroke and transient ischemic attack patients: A joint statement by the American Society of Neuroradiology, the American College of Radiology, and the Society of NeuroInterventional Surgery. *AJNR. American Journal of Neuroradiology.* 2013; 34: E117–E127. <https://doi.org/10.3174/ajnr.A3690>.
- [8] Mokli Y, Pfaff J, Dos Santos DP, Herweh C, Nagel S. Computer-aided imaging analysis in acute ischemic stroke - background and clinical applications. *Neurological Research and Practice.* 2019; 1: 23. <https://doi.org/10.1186/s42466-019-0028-y>.

- [9] Daidone M, Ferrantelli S, Tuttolomondo A. Machine learning applications in stroke medicine: advancements, challenges, and future perspectives. *Neural Regeneration Research*. 2024; 19: 769–773. <https://doi.org/10.4103/1673-5374.382228>.
- [10] Ali F, Hamid U, Zaidat O, Bhatti D, Kalia JS. Role of Artificial Intelligence in TeleStroke: An Overview. *Frontiers in Neurology*. 2020; 11: 559322. <https://doi.org/10.3389/fneur.2020.559322>.
- [11] Liu Y, Shah P, Yu Y, Horsey J, Ouyang J, Jiang B, *et al*. A Clinical and Imaging Fused Deep Learning Model Matches Expert Clinician Prediction of 90-Day Stroke Outcomes. *AJNR. American Journal of Neuroradiology*. 2024; 45: 406–411. <https://doi.org/10.3174/ajnr.A8140>.
- [12] Jiang B, Pham N, van Staalduinen EK, Liu Y, Nazari-Farsani S, Sanaat A, *et al*. Deep Learning Applications in Imaging of Acute Ischemic Stroke: A Systematic Review and Narrative Summary. *Radiology*. 2025; 315: e240775. <https://doi.org/10.1148/radiol.240775>.
- [13] Wei L, Pan X, Deng W, Chen L, Xi Q, Liu M, *et al*. Predicting long-term outcomes for acute ischemic stroke using multi-model MRI radiomics and clinical variables. *Frontiers in Medicine*. 2024; 11: 1328073. <https://doi.org/10.3389/fmed.2024.1328073>.
- [14] Krag CH, Müller FC, Gandrup KL, Raaschou H, Andersen MB, Brejnbøl MW, *et al*. Diagnostic test accuracy study of a commercially available deep learning algorithm for ischemic lesion detection on brain MRIs in suspected stroke patients from a non-comprehensive stroke center. *European Journal of Radiology*. 2023; 168: 111126. <https://doi.org/10.1016/j.ejrad.2023.111126>.
- [15] Sirsat MS, Fermé E, Câmara J. Machine Learning for Brain Stroke: A Review. *Journal of Stroke and Cerebrovascular Diseases: the Official Journal of National Stroke Association*. 2020; 29: 105162. <https://doi.org/10.1016/j.jstrokecerebrovasdis.2020.105162>.
- [16] Liu Y, Wen Z, Wang Y, Zhong Y, Wang J, Hu Y, *et al*. Artificial intelligence in ischemic stroke images: current applications and future directions. *Frontiers in Neurology*. 2024; 15: 1418060. <https://doi.org/10.3389/fneur.2024.1418060>.
- [17] Goyal M, Demchuk AM, Menon BK, Eesa M, Rempel JL, Thornton J, *et al*. Randomized assessment of rapid endovascular treatment of ischemic stroke. *The New England Journal of Medicine*. 2015; 372: 1019–1030. <https://doi.org/10.1056/NEJMoa1414905>.
- [18] Campbell BCV, Mitchell PJ, Kleinig TJ, Dewey HM, Churilov L, Yassi N, *et al*. Endovascular therapy for ischemic stroke with perfusion-imaging selection. *The New England Journal of Medicine*. 2015; 372: 1009–1018. <https://doi.org/10.1056/NEJMoa1414792>.
- [19] Albers GW, Marks MP, Kemp S, Christensen S, Tsai JP, Ortega-Gutierrez S, *et al*. Thrombectomy for Stroke at 6 to 16 Hours with Selection by Perfusion Imaging. *The New England Journal of Medicine*. 2018; 378: 708–718. <https://doi.org/10.1056/NEJMoa1713973>.
- [20] Lai WC, Guo L, Bialkowski K, Bialkowski A. An Explainable Deep Learning Method for Microwave Head Stroke Localization. *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*. 2023; 7: 336–343. <https://doi.org/10.1109/JERM.2023.3287681>.
- [21] Limam H, Jouini E. Comparative Approach of Deep Learning Methods for Predicting Functional Outcomes After Stroke. *International Conference on Optimization and Data Science in Industrial Engineering* (pp. 137–149). Springer: Cham. 2026. [https://doi.org/10.1007/978-3-031-93601-2\\_9](https://doi.org/10.1007/978-3-031-93601-2_9).
- [22] Tuxunjiang P, Huang C, Zhou Z, Zhao W, Han B, Tan W, *et al*. Prediction of NIHSS Scores and Acute Ischemic Stroke Severity Using a Cross-attention Vision Transformer Model with Multimodal MRI. *Academic Radiology*. 2025; 32: 5453–5467. <https://doi.org/10.1016/j.acra.2025.05.031>.