Development and Validation of a Diffusion-weighted Imaging-only Deep Learning Model for the Detection of Supratentorial Acute Ischemic Infarcts: A Retrospective Diagnostic Accuracy Study

Saketh Kumar Katam, Shivanand Patil, Suresh Kanamadi

Department of Radio-Diagnosis, Shri B.M. Patil Medical College, Hospital and Research Centre and BLDE (Deemed to be University), Vijayapura, Karnataka, India

Abstract

Background: Acute ischemic stroke is a leading cause of death and long-term disability worldwide, with a disproportionately increasing burden in low- and middle-income countries. Diffusion-weighted imaging (DWI) is the most sensitive magnetic resonance imaging (MRI) sequence for the early detection of acute ischemia, as treatment decisions, including intravenous thrombolysis and mechanical thrombectomy, are highly time-dependent. However, interpretation of subtle lesions may be challenging, particularly in high-volume emergency settings, potentially leading to diagnostic delays. **Objective:** To train and test a DWI-based deep learning model for the automated detection of supratentorial acute ischemic infarcts and assess its diagnostic performance relative to expert radiologist interpretation. **Materials and Methods:** This retrospective diagnostic accuracy study included adult patients who were suspected of having acute ischemic stroke and underwent an MRI the brain in a tertiary care teaching hospital. Only the supratentorial DWI images were analyzed. The studies were classified into normal or infarct-positive according to consensus interpretation of the experienced radiologists with correlation to apparent diffusion coefficient (ADC) maps. The deep learning-based classification model with 70:15:15 train–validation–test split was trained and tested using images that had undergone preprocessing. Model performance was assessed using accuracy, sensitivity, specificity, F1-score, and area under the curve (AUC). **Results:** A total of 1024 MRI studies were included. The proposed model showed high diagnostic accuracy on the independent test data, with an overall accuracy of 90.8%, a sensitivity of 92.1%, a specificity of 89.6%, and an AUC of 0.93. The average processing time per study was <1 s. **Conclusion:** The DWI-based deep learning model demonstrated high diagnostic accuracy in the identification of supratentorial acute ischemic infarcts. This system may serve as a decision-support tool to assist radiologists and enhance workflow efficiency, particularly in time-sensitive and resource-constrained settings.

Keywords: Acute ischemic stroke, artificial intelligence, deep learning, diagnostic accuracy, diffusion-weighted imaging

Résumé

Contexte: L'accident vasculaire cérébral ischémique aigu constitue une cause majeure de mortalité et d'invalidité à long terme dans le monde, avec une charge croissante dans les pays à revenu faible et intermédiaire. L'imagerie par résonance magnétique pondérée en diffusion (DWI) est la séquence la plus sensible pour la détection précoce, mais l'interprétation des lésions subtiles peut être difficile en pratique clinique, en particulier dans les contextes à forte charge de travail. **Objectif:** Développer et valider un modèle d'apprentissage profond basé sur la DWI pour la détection automatisée des infarctus ischémiques aigus supratentoriels et évaluer ses performances diagnostiques par rapport à l'interprétation des radiologues experts. **Matériels et Méthodes:** Cette étude rétrospective de précision diagnostique a inclus des patients adultes suspects d'AVC ischémique aigu ayant bénéficié d'une IRM cérébrale. Seules les images DWI supratentorielles ont été analysées. Les cas ont été classés sur la base d'un consensus entre radiologues expérimentés avec

Address for correspondence: Dr. Shivanand Patil,
Department of Radio-Diagnosis, Shri B.M. Patil Medical College,
Hospital and Research Centre and BLDE (Deemed to be University),
Vijayapura, Karnataka, India.
E-mail: drshivbjp@gmail.com

Access this article online

Quick Response Code:

Website: www.annalsafrmed.org

DOI: 10.4103/aam.aam_224_26

How to cite this article: Katam SK, Patil S, Kanamadi S. Development and validation of a diffusion-weighted imaging-only deep learning model for the detection of supratentorial acute ischemic infarcts: A retrospective diagnostic accuracy study. Ann Afr Med 0;0:0.

Submitted: 01-Mar-2026 **Revised:** 01-Mar-2026
Accepted: 24-Mar-2026 **Published:** 14-May-2026

corrélation aux cartes ADC. Un modèle d'apprentissage profond a été entraîné selon une répartition 70:15:15 (entraînement–validation–test). Les performances ont été évaluées à l'aide de l'exactitude, de la sensibilité, de la spécificité, du score F1 et de l'aire sous la courbe (AUC). Résultats: Un total de 1024 examens IRM a été inclus. Le modèle a montré une exactitude de 90,8 %, une sensibilité de 92,1 %, une spécificité de 89,6 % et une AUC de 0,93 sur l'ensemble de test indépendant. Le temps moyen de traitement par examen était inférieur à une seconde. **Conclusion:** Le modèle d'apprentissage profond basé sur la DWI a démontré une haute précision diagnostique pour la détection des infarctus ischémiques aigus supratentoriels et pourrait servir d'outil d'aide à la décision pour améliorer l'efficacité du flux de travail en radiologie, en particulier dans les contextes sensibles au temps et à ressources limitées.

Mots-clés: Accident vasculaire cérébral ischémique aigu, apprentissage profond, imagerie pondérée en diffusion, intelligence artificielle, précision diagnostique

Introduction

Stroke remains a major global health challenge and is among the leading causes of mortality and long-term disability worldwide.[1,2] Approximately 80%–85% of all strokes are ischemic and may result in significant morbidity if not diagnosed and treated promptly (a concept widely summarized as "time is brain").[3-6]

Neuroimaging is a central factor in suspected acute stroke assessment. While noncontrast computed tomography is widely used to exclude hemorrhage, its sensitivity for early ischemic changes in the hyperacute phase is limited. Magnetic resonance imaging (MRI), particularly diffusion-weighted imaging (DWI), is the most sensitive modality for the early detection of acute ischemic infarction.[7,8] DWI identifies cytotoxic edema as hyperintensity with corresponding apparent diffusion coefficient (ADC) reduction, often within minutes of symptom onset.[9-13]

Despite its diagnostic value, the interpretation of DWI can be challenging. Assessment may be complicated by motion artifacts, susceptibility effects, small cortical and lacunar infarcts, and T2 shine-through events.[11] Diagnostic accuracy may also be impacted by radiologist experience variability and reporting delays in congested emergency rooms. These issues are especially pertinent in healthcare systems when support for specialized neuroradiology is limited.

Artificial intelligence (AI) and image analysis based on deep learning, in particular, have been shown to be effective at helping with the interpretation of medical images, and in stroke imaging specifically, lesion detection and segmentation have been demonstrated to be as effective as a human reader.[14-17] AI systems have had some positive results to date, but most of these prior systems are dependent on curated datasets, so they may not be as practical to use in resource-deprived environments.

A simplified, single-sequence DWI-based model trained on real-world institutional data may offer a scalable and clinically practical solution, particularly in resource-constrained healthcare environments. The present study aimed to develop and validate such a model and evaluate its diagnostic performance against expert radiologist interpretation.

Materials and Methods

Study design and ethics

This retrospective observational diagnostic accuracy study was conducted in the department of radiodiagnosis at a tertiary care teaching hospital in southern India. The Institutional Ethics Committee approval was obtained (IEC No. BLDE (DU)/IEC-SBMPMC/037/2023-24 dated February 10, 2024). All imaging data were anonymized before analysis. Informed consent was waived due to the retrospective use of deidentified data.

Study population

Adult patients (≥18 years) who underwent an MRI of the brain for suspected acute ischemic stroke during the study period were eligible. Only studies with adequate image quality and clear visualization of supratentorial structures were included.

Inclusion criteria

MRI of the brain studies with DWI (b = 1000 s/mm^2) demonstrating either normal supratentorial findings or acute supratentorial ischemic infarction confirmed on ADC maps. The time from symptom onset to MRI acquisition was not used as an inclusion criterion.

Exclusion criteria

Motion-degraded or artifact-dominant studies, posterior fossa-only infarcts, nonischemic causes of DWI hyperintensity such as tumors, abscesses, or encephalitis, postoperative cases, and pediatric patients.

All eligible cases during the study period were included; no formal sample size calculation was performed. By the end of the study, a total of 1024 studies were included in model training, validation, and testing.

Reference standard

Two radiologists with over 10 years of neuroimaging experience independently assessed each DWI study. Consensus was used to settle disagreements. In every instance, ADC maps were examined to rule out T2 shine-through and verify actual restricted diffusion. The reference standard for segmentation analysis was manual lesion annotations.

Image processing and capturing

The MRI examinations were performed on a 1.5 Tesla scanner (GE SIGNA Explorer, GE Healthcare) using a standard acute stroke protocol. DWI was acquired with a b value of 1000 s/mm^2.

DWI images were retrieved from the institutional PACS and converted from DICOM format into numerical arrays. Only supratentorial slices were retained. Preprocessing included intensity normalization, resizing to uniform spatial resolution, and data augmentation to improve model robustness.

Model architecture

A multiscale convolutional neural network architecture (MUSCLE-Net) was developed to capture both fine-grained and contextual diffusion signal features. The implementation of the model was in Python through TensorFlow. To improve generalization, conventional regularization methods were used during training.

Postprocessing included atlas-based regional probability mapping using academic software developed at Johns Hopkins University (© 2022), used under the JHU Academic Software License.

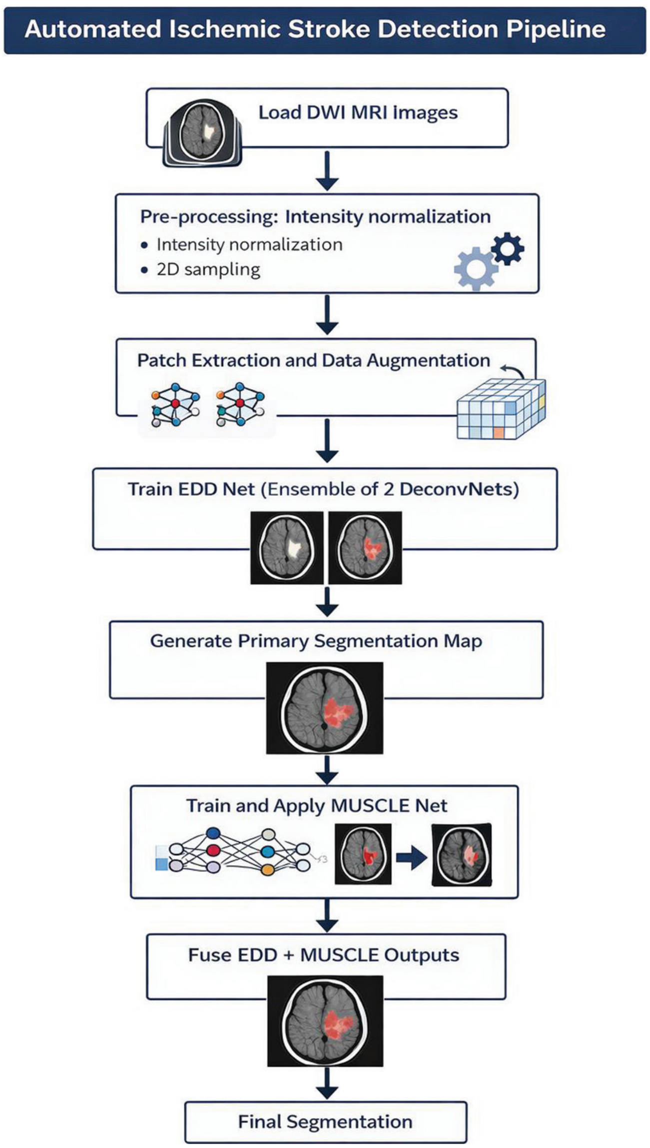

Figure 1: Overview of the proposed deep learning framework. Diffusion-weighted images undergo preprocessing (intensity normalization and two-dimensional sampling), followed by patch extraction and augmentation. An encoder–decoder deconvolutional network (EDD Net) generates a primary segmentation map, which is refined using MUSCLE-Net. Fused outputs yield the final infarct segmentation

The overall architecture and workflow of the proposed framework are illustrated in Figure 1.

Model training and evaluation

The dataset was randomly split into training (70%), validation (15%), and testing (15) subsets to prevent data leakage between sets [Table 1]. Training of the model was done with the Adam optimizer and binary cross-entropy loss. Early stopping was implemented based on validation performance to prevent overfitting.

Training was done with input images that were standardized to uniform spatial resolution. Random rotations, flipping, and noise addition techniques were used as data augmentation techniques to enhance robustness.

The independent test set was assessed based on accuracy, sensitivity, specificity, precision, F1-score, confusion matrix analysis, and receiver operating characteristic (ROC) curve analysis, with the calculation of the area under the curve (AUC). Diagnostic accuracy metrics and confidence intervals of 95% were determined with the help of the Wilson score method, and AUC confidence intervals were determined with the help of the Hanley–McNeil method. The Dice similarity coefficient was used to measure segmentation performance.

Results

Overall model performance

The model demonstrated high diagnostic accuracy in distinguishing normal from infarct positive supratentorial DWI studies on the independent test dataset [Table 2].

ROC analysis demonstrated high separability between normal and infarct-positive cases across probability thresholds (AUC = 0.93), as shown in Figure 2.

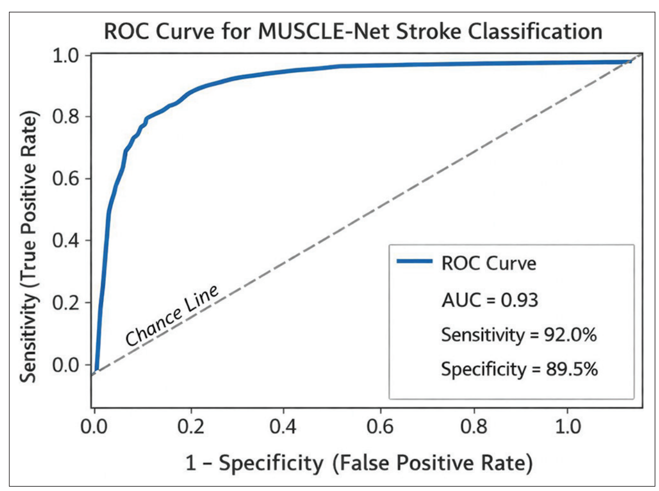

Figure 2: Receiver operating characteristic curve demonstrating classification performance of the MUSCLE-Net model for detection of supratentorial acute ischemic infarcts on diffusion-weighted magnetic resonance imaging. The model achieved an area under the curve of 0.93, with a sensitivity of 92% and specificity of 89.5% at the selected threshold. The dashed line represents chance performance

Table 1: Dataset split

Subset	Number of scans	Normal	Infarct
Training set	717	358	359
Validation set	154	77	77
Test set	153	77	76

Table 2: Overall performance

Metric	Result	95% CI	Remarks
Mean Dice similarity coefficient	0.65	0.61–0.69	Segmentation performance
Accuracy (%)	90.8	86.3–95.4	Overall performance
Sensitivity (%)	92.1	86–98.2	Reflects the ability to detect infarcts
Specificity (%)	89.6	82.8–96.4	Reflects the ability to exclude normal scans
AUC	0.93	0.89–0.97	ROC analysis (discriminative performance)

CI=Confidence interval, AUC=Area under the curve, ROC=Receiver operating characteristic

Segmentation-based performance evaluation

Besides the classification at the study level, segmentation-based evaluation was conducted to evaluate the spatial agreement between automated results and the expert manual annotations. The mean Dice similarity coefficient of the dataset was 0.65, indicating substantial spatial agreement between predicted and reference segmentations. The observed value of 0.65 was consistent with previously reported benchmarks in similar DWI-based segmentation studies, which typically ranged between 0.50 and 0.75, depending on lesion size and acquisition parameters.

The performance of segmentation was observed to be variable with regard to lesion size and morphology. Small cortical and lacunar infarcts, in contrast, exhibited relatively lower Dice values, probably as a result of partial volume effects, subtle diffusion signal alterations, and susceptibility-related distortions inherent to echo-planar DWI acquisition.

Lesion size-based analysis

To evaluate robustness further, infarcts were classified as small and large lesions according to the approximate lesion volume. Both the groups had a high detection performance, and a slight decrease in segmentation accuracy was found in smaller lesions [Table 3].

Small infarcts represented a significant percentage of the dataset and were the most challenging to deal with because they have a small spatial range and a lower contrast-to-noise ratio. However, the model showed a consistent identification of clinically relevant lesions with few cases of false-negativity in very small or punctate infarcts.

False positive and false negative analysis

The error analysis was done in detail to describe false-positive and false-negative predictions. T2 shine-through effects, susceptibility artifacts around the skull base and air–bone interfaces, and periventricular signal heterogeneity were the most related to false-positive detections. The hybrid model that included refinement steps greatly minimized these spurious detections by quenching isolated or artifactual spots that do not align with ischemic diffusion patterns. There were a few cases of false-negative, mostly due to very small cortical infarcts and deep lacunar infarcts with insignificant diffusion restriction. No trend in bias, in terms of lesion laterality or vascular territory, was found.

Qualitative visual assessment

Qualitative analysis showed good visual correspondence between AI-generated results and manual annotation by experts. Representative examples are provided of original DWI images and the results of AI-generated outputs, which demonstrate precise localization of ischemic lesions of different sizes and locations.

The model was able to detect infarcts in cortical, subcortical, and deep gray matter areas and reduced spurious detection of infarcts in normal brain parenchyma. The visual inspection revealed that refinement steps were very important in enhancing interpretability and minimizing artifactual outputs.

Representative qualitative examples demonstrating correspondence between original DWI images, expert annotations, and AI-generated outputs are shown in Figure 3.

Processing time analysis

The mean time of processing per MRI study was below a second, which reflects high computational efficiency. This fast inference time makes the integration of the proposed framework in the normal clinical practice a possibility, especially in emergency stroke imaging, where fast triage and reporting are needed.

Discussion

The proposed model demonstrated strong discriminative performance for automated detection of supratentorial acute ischemic infarcts on DWI and is comparable to previously reported AI-based stroke detection systems. Clinically, high sensitivity is desirable because missed infarcts may delay reperfusion therapy, whereas the specificity attained implies successful differentiation between actual ischemic lesions and normal supratentorial DWI signal changes. Segmentation analysis yielded a mean Dice similarity coefficient of 0.65, indicating substantial spatial agreement with expert manual annotations.

Predictably, the performance of segmentation was better with larger infarcts with well-defined margins than with smaller cortical and lacunar infarcts. This drawback is in line with the previous research and can be explained largely by partial volume effects, subtle diffusion constraints, and susceptibility-related artifacts inherent to DWI rather than an inherent limitation of the model architecture.

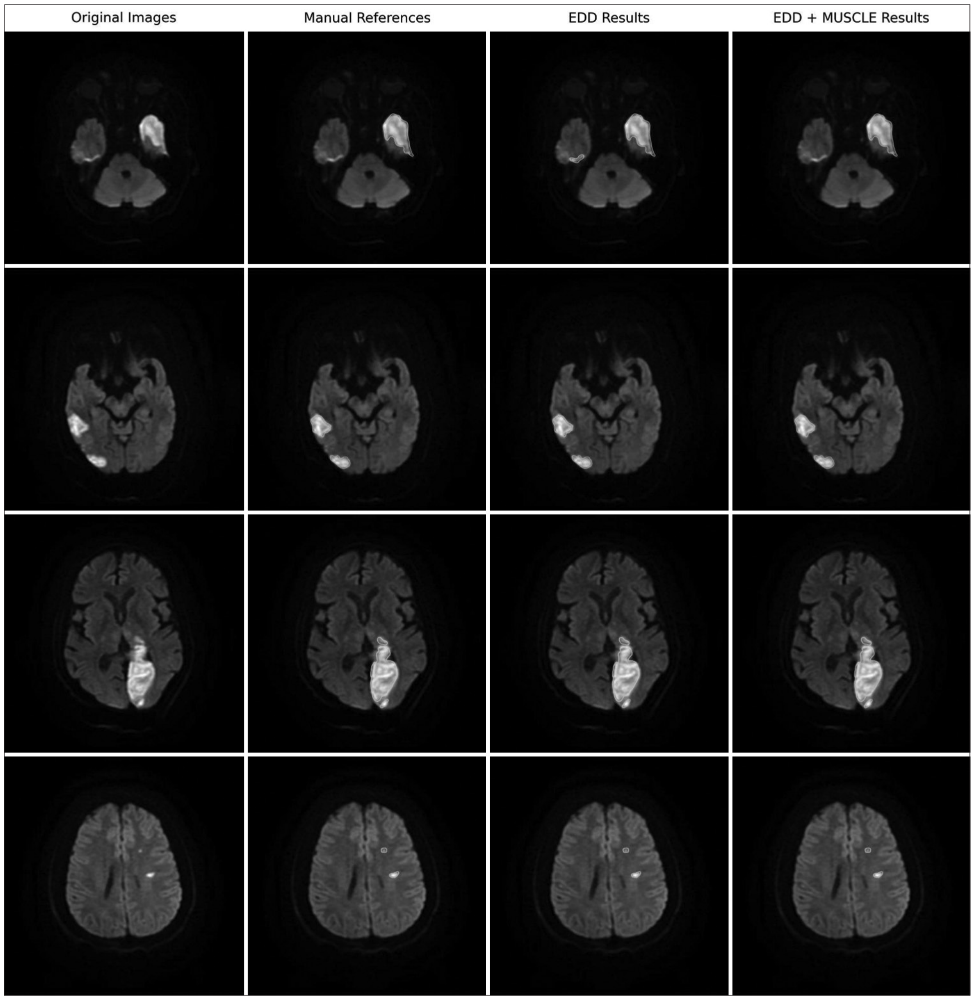

Figure 3: Representative axial diffusion-weighted magnetic resonance imaging (diffusion-weighted imaging [DWI]) images illustrating lesion segmentation performance. The first column displays the original DWI input images. The second column shows manual reference annotations outlined in yellow. The third column shows automated segmentation results produced by the encoder–decoder (EDD) model. The fourth column shows automated segmentation results produced by the combined EDD + MUSCLE model. Yellow contours indicate the segmented lesion regions

Table 3: Small versus large lesions

Lesion category	Proportion (%)	Mean Dice coefficient	Detection rate	Remarks
Small lesions	68	0.59	0.91	Lower Dice due to subtle signal
Large lesions	32	0.80	0.99	Better lesion delineation

The use of a multiscale convolutional learning architecture is a major strength of the study. MUSCLE-Net can extract both fine and contextual features of the image simultaneously, which allows it to detect lesions of different sizes and locations. The strategy of refinement implemented in the framework had further reduced false-positive detections because of T2 shine-through and artifactual signal differences, enhancing clinical interpretability.

Clinical and public health implications

There are significant practical benefits associated with the exclusive use of DWI as the model input. DWI is the key to every acute stroke MRI protocol, has minimal acquisition time, and is highly lesion conspicuous during the hyperacute phase. Single-sequence methodology eases preprocessing, lowers computational expenses, and increases the possibility of implementation in high-volume and resource-constrained clinical settings.[18-22] Moreover, the training and assessment of institutional data based on an Indian patient population enhances practicality and possible external validity in other similar healthcare environments.[23-25]

Overall, the results suggest that a deep learning framework based on DWI can be used to assist radiologists in acute stroke imaging by enhancing diagnostic confidence, decreasing the variability, and helping them make timely clinical decisions.

Limitations

There are several limitations to this study. First, it was retrospective in design and carried out in a single tertiary care center, which might restrict the extrapolation of the results in other institutions, scanner manufacturers, and imaging regimes. Second, only DWI was used to develop the model, but again, this makes it more feasible and easier to integrate workflows; further integration of other MRI sequences, like FLAIR or perfusion imaging, could contribute to better diagnostic performance.

Very small infarcts, especially lacunar and punctate cortical lesions, were less accurately segmented, mainly because of intrinsic diffusion imaging constraints, such as partial volume effects and susceptibility artifacts. Moreover, the dataset was not available publicly because of institutional and patient privacy restrictions, which might restrict independent external validation. Even though the time of inference was very low, the model had been tested in an offline research setting, and future validation using real-time PACS integration was not within the scope of this study.

Conclusion

A DWI-only deep learning model demonstrated high diagnostic accuracy for the detection of supratentorial acute ischemic infarcts on MRI. The proposed framework offers a practical and computationally efficient approach that may be suitable for integration into routine acute stroke workflows, particularly in time-sensitive and resource-limited clinical settings.

Although not intended to replace radiologists, such an AI-based system may serve as a decision-support tool to enhance diagnostic confidence and workflow efficiency. Further prospective multicenter validation and external testing are warranted before routine clinical implementation.

Data availability statement

The imaging data used in this study are not publicly available due to institutional data protection policies and patient privacy regulations. However, anonymized data may be made available from the corresponding author upon reasonable request and subject to approval by the Institutional Ethics Committee. The source code for model development, training, and evaluation will be made available to academic researchers upon reasonable request to facilitate reproducibility and independent validation.

Acknowledgments

The authors acknowledge the use of academic software components developed by Andreia V. Faria and Chin-Fu Liu, Johns Hopkins University (© 2022), distributed via the Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC) and used under the JHU Academic Software License for atlas-based regional feature extraction and postprocessing in this study.[26]

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Feigin VL, Norrving B, Mensah GA. Global burden of stroke. Lancet. 2014;383:245-55.
2. Donnan GA, Fisher M, Macleod M, Davis SM. Stroke. Lancet 2008;371:1612-23.
3. Saver JL. Time is brain – Quantified. Stroke 2006;37:263-6.
4. Pandian JD, Sudhan P. Stroke epidemiology and stroke care services in India. J Stroke 2013;15:128-34.
5. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, *et al.* Heart disease and stroke statistics-2016 update. Circulation 2016;133:e38-360.
6. Astrup J, Siesjö BK, Symon L. Thresholds in cerebral ischemia – The ischemic penumbra. Stroke 1981;12:723-5.
7. Baird AE, Warach S. Magnetic resonance imaging of acute stroke. J Cereb Blood Flow Metab 1998;18:583-609.
8. Warach S, Chien D, Li W, Ronthal M, Edelman RR. Fast magnetic resonance diffusion-weighted imaging of acute human stroke. Neurology 1992;42:1717-23.
9. Stejskal EO, Tanner JE. Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. J Chem Phys 1965;42:288-92.
10. Le Bihan D, Mangin JF, Poupon C, Clark CA, Pappata S, Molko N, *et al.* Diffusion tensor imaging: Concepts and applications. J Magn Reson Imaging 2001;13:534-46.
11. Lutsep HL, Albers GW, DeCrespigny A, Kamat GN, Marks MP, Moseley ME. Clinical utility of diffusion-weighted magnetic resonance imaging in the assessment of ischemic stroke. Ann Neurol 1997;41:574-80.
12. Kidwell CS, Chalela JA, Saver JL, Starkman S, Hill MD, Demchuk AM, *et al.* Comparison of MRI and CT for detection of acute intracerebral hemorrhage. Stroke. 2004;35:505-10.
13. Barber PA, Darby DG, Desmond PM, Yang Q, Gerraty RP, Jolley D, *et al.* Prediction of stroke outcome with echoplanar perfusion- and diffusion-weighted MRI. Neurology 1998;51:418-26.
14. Litjens G, Kooi T, Bejnordi BE, Setio AA, Ciompi F, Ghafoorian M, *et al*. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60-88.
15. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, *et al.* Dermatologist-level classification of skin cancer with deep neural networks. Nat Med 2017;23:24-30.
16. Chen L, Bentley P, Rueckert D. Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks. Neuroimage Clin 2017;15:633-43.
17. Maier O, Menze BH, von der Gablentz J, Ȟani L, Heinrich MP, Liebrand M, *et al.* ISLES 2015 – A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. Med Image Anal 2017;35:250-69.
18. Kamnitsas K, Ledig C, Newcombe VF, Simpson JP, Kane AD, Menon DK, *et al.* Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med Image Anal 2017;36:61-78.
19. Soun JE, Chow DS, Nagamine M, Takhtawala RS, Filippi CG, Yu W, *et al*. Artificial intelligence and acute stroke imaging. Radiology. 2021;298:6-17.
20. Nielsen A, Hansen MB, Tietze A, Mouridsen K. Prediction of tissue outcome and assessment of treatment effect in acute ischemic stroke using deep learning. Stroke 2018;49:1394-401.
21. Offersen N, Hansen MB, Mouridsen K. Deep learning-based assessment of diffusion-FLAIR mismatch in acute ischemic stroke. Stroke. 2023;54:e45–54.
22. Zijdenbos AP, Dawant BM, Margolin RA, Palmer AC. Morphometric analysis of white matter lesions in MR images: Method and validation. IEEE Trans Med Imaging 1994;13:716-24.
23. Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL, *et al.* Open access series of imaging studies (OASIS): Cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. NeuroImage. 2007;33:594-604.
24. Heit JJ, Wintermark M. Perfusion computed tomography for the evaluation of acute ischemic stroke. Neuroimaging Clin N Am.

2018;28:465-77.
25. Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. Nat Med 2019;25:44-56.
26. Faria AV, Liu CF. Quantitative Flow Visualization (QFV) and Regional Feature Extraction Software for Stroke Imaging. Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC); 2022. Available from: https://www.nitrc.org/frs/?group_id=1520. [Last accessed on 2026 Jan 27].