International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 21, Number 3 (2026) pp. 401-414
Research India Publications
https://www.ripublication.com

Braintem: A Cloud-Based Decision-Support Platform For Dwi-Mri Stroke Lesion Analysis

Izzatul Husna Azman[1], Abdul Rahim Abdullah[1]*, Norhashimah Mohd Saad[2], Ezreen Farina Shair[1], Ahmad Sobri Muda[3] and Farzanah Atikah Yamba[4]

[1] *Faculty of Electrical Technology and Engineering, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia*

[2] *Faculty of Electronic Technology and Engineering, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia*

[3] *Faculty of Medicine and Health Sciences, Universiti Putra Malaysia (UPM), 43400 Serdang, Malaysia*

[4] *PadiMedical Division, Lönge Medikal Sdn Bhd, Universiti Putra Malaysia (UPM), 43400 Serdang, Malaysia*

***Correspondence: Abdul Rahim Abdullah*

Abstract

Diffusion-weighted magnetic resonance imaging (DWI-MRI) supports assessment of acute ischemic stroke, yet integrating image analysis into an accessible workflow remains challenging. This paper presents BrainTeM, a browser-based research prototype that combines DWI preparation, candidate-lesion segmentation, feature extraction, classification, and visual reporting. The pipeline applies intensity normalization, background removal, and gamma enhancement, followed by fuzzy c-means and active contours (FCMAC). Thirteen measurements are supplied to a weighted k-nearest-neighbour classifier. Users upload DICOM images or ZIP packages and review processed images, segmentation overlays, quantitative outputs, and predicted categories with confidence values. The project collection comprises 166 cases from ISLES 2015, Malaysian hospitals, and PadiMedical/Lönge Medikal. The current workflow analyzes selected annotated slices, not complete volumes. The deployed model record reports 93.10% test accuracy, 93.86% cross-validation accuracy, and 65.08% balanced accuracy. Because split details, test-set size, and class-wise results were unavailable, these scores are preliminary and do not establish clinical performance. BrainTeM supports research review, not diagnosis; independent patient-level validation, usability testing, and governance review remain necessary.

Keywords: BrainTeM; diffusion-weighted MRI; ischemic stroke; lesion segmentation; machine learning; cloud application.

Introduction

Stroke is a time-critical neurological condition, and imaging plays an important role in identifying and characterizing lesions and informing treatment planning. In acute ischemia, diffusion-weighted imaging (DWI) can reveal restricted water diffusion early after symptom onset [1][2]. Lesion appearance nevertheless varies with size, location, acquisition settings, artifacts, and time from onset. Manual delineation can be demanding, while comparisons between computational methods are complicated by differences in datasets, annotations, and validation protocols [3]. Recent global estimates continue to show a substantial stroke burden [4][5]. Figure 1 summarizes the motivation for DWI-based analysis and structured human review [6 - 8]. Recent work relevant to this study includes an earlier review of MRI-based computer-aided diagnosis (CAD) for stroke [9] and our subsequent empirical DWI-MRI CAD study using FCMAC segmentation and machine-learning classifiers [10]. For more recent benchmarking, ISLES 2022 provides a multicentre stroke-lesion dataset [11]. DeepISLES reports evaluation on an external real-world cohort [12], while a 2026 study presents an adaptive deep-learning method for DWI lesion segmentation [13].

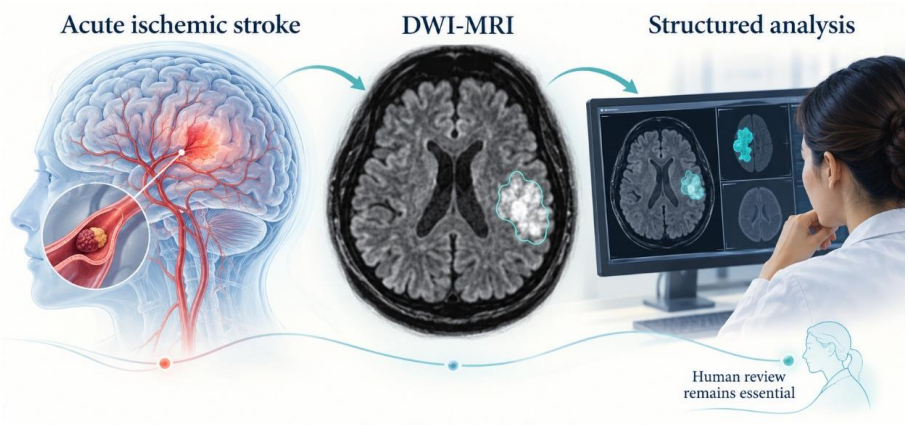


Figure 1. Conceptual context for DWI-based stroke lesion analysis and structured result review [1-5,9].

Automated stroke-lesion analysis has progressed from classical image processing and machine learning to deep-learning methods. ISLES 2015 offered a shared benchmark for multispectral MRI segmentation, while also illustrating how heterogeneous cases make algorithm comparisons difficult [3]. Earlier DWI studies explored fuzzy clustering, active contours, engineered features, and supervised classification [6]. Convolutional models have since shown promise, although small or heterogeneous datasets, domain shift, and limited external validation remain concerns [7–8]. A 2025 systematic review likewise found substantial variation across studies and limited generalisability in acute and subacute MRI lesion segmentation [14].

BrainTeM brings an existing DWI analysis pipeline into a web-based cloud workflow. The cloud workflow builds on our earlier empirical DWI-MRI CAD study, which evaluated the FCMAC and machine-learning pipeline using the project data [10]. The present paper focuses on the browser-accessible deployment and structured result-review workflow; any reused cohort or model outputs should be identified as

previously reported rather than presented as new findings. It links image submission with preprocessing, candidate lesion segmentation, measurement, classification, and result review. This paper examines the system integration and its operation. It does not introduce a new segmentation algorithm or claim clinical validation or improved patient outcomes.

This paper highlights three contributions. It describes how the existing FCMAC pipeline and 13-feature weighted k-nearest-neighbour classifier were integrated into a browser-based application; outlines the cloud arrangement separating the interface, analysis service, model files, and result storage; and reports the model-summary values currently available, together with their evaluation limitations and the validation still needed before clinical use.

Literature Review

DWI-based ischemic stroke analysis

DWI is sensitive to changes in water diffusion and is routinely used when assessing acute ischemic lesions [1–2]. Automated segmentation remains difficult because lesions can be small, irregular, or hard to distinguish from nearby structures. ISLES 2015 provided a public dataset and benchmark for comparing stroke-lesion segmentation methods [3]. Results on one benchmark, however, do not establish performance across scanners, hospitals, acquisition protocols, or stages of disease.

2.2. Classical segmentation and machine learning

Fuzzy c-means (FCM) gives each pixel or feature vector graded membership in several clusters, rather than assigning it immediately to a single group [15]. This can help represent uncertain tissue boundaries, although standard FCM is sensitive to initialization and noise. Active-contour methods refine a boundary by evolving a curve toward image structures while minimizing an energy function [16]. In a hybrid approach, FCM membership can help initialize or guide that evolution. For a small, curated dataset, a feature-based classifier is a compact option compared with training a high-capacity neural network. Weighted k-nearest neighbours (WKNN) use nearby training samples to assign a class, with closer samples having greater influence [17].

Earlier DWI studies have combined FCM, active contours, statistical features, and machine-learning classifiers for stroke analysis [6]. BrainTeM follows this line of work and makes the pipeline available through a browser. Accordingly, this paper concerns the delivery and integration of the system; it does not present FCMAC or WKNN as new methods.

2.3. From algorithm prototype to usable application

A medical-imaging application involves more than its model score. Implementation also must handle image formats and metadata, storage, user access, traceability, and review of outputs. DICOM standardizes the communication and management of medical-imaging information, although files can still differ in encoding and series structure [18]. A browser interface can spare users from installing analysis software, and a cloud service can provide a common execution environment. Neither feature demonstrates that a system is secure, clinically validated, or suitable for every institution; those questions require testing and governance.

Moving AI into healthcare raises questions about dataset shift, workflow fit, evaluation design, and accountability [19]. Reported performance can also be overstated when studies include data leakage, non-independent splits, or incomplete reporting [8,20]. A research prototype should therefore present its output as aids for review, communicate uncertainty, and undergo independent validation before it can inform patient care.

2.4. Implementation, governance, and human factors

A usable image analysis service must fit the maturity of its intended workflow, with clear responsibilities and users. A radiology workflow roadmap distinguishes research systems from production integrations and feedback-enabled systems, each of which has different requirements [21-22]. BrainTeM is presented here as a research-support prototype. It has not been integrated with a hospital PACS or validated as part of a clinical pathway.

Health-AI governance frameworks emphasize human oversight, privacy, safety, accountability, and equitable performance [23-24]. Guidance for medical-device development also stresses lifecycle controls, representative data, monitoring, and a clearly defined intended use [25-26]. Consistent with these principles, BrainTeM presents outputs for review and does not authorize clinical action.

2.5. Evaluation and reporting of medical AI

Technical performance and the effect of decision-support software on users or care processes are separate questions. Guidance on clinical evaluation recommends appropriate cohorts, comparators, and statistical methods [27]. CONSORT-AI and SPIRIT-AI call for clear reporting of AI interventions, input data, workflow, and human-AI interaction [28-29]. TRIPOD+AI and DECIDE-AI provide further guidance for reporting prediction models and evaluating early-stage decision-support systems [30-31].

Image overlays and confidence values make intermediate outputs easier to inspect, but they do not explain by themselves why a model produced a prediction. Explainability in healthcare involves several disciplines, and current techniques have limitations that should be stated clearly [32-33]. The quality of human-AI collaboration must therefore be evaluated in the workflow itself; it cannot be inferred from the presence of an interface [34].

Materials and Methods

System design

BrainTeM is a browser-accessible application built around a Flask service deployed on Google Cloud Run. It accepts DICOM files or ZIP packages, starts the image-analysis pipeline, and returns a job status with a visual result bundle. Firebase services support user identity, job metadata, and object storage in the project architecture. The design separates public information pages, the authenticated analysis portal, backend processing, and stored results so each component can be managed independently. Figure 2 outlines the data flow. As in other research-service patterns, an analysis service receives inputs, generates outputs, and returns them to a browser for inspection [21][22].

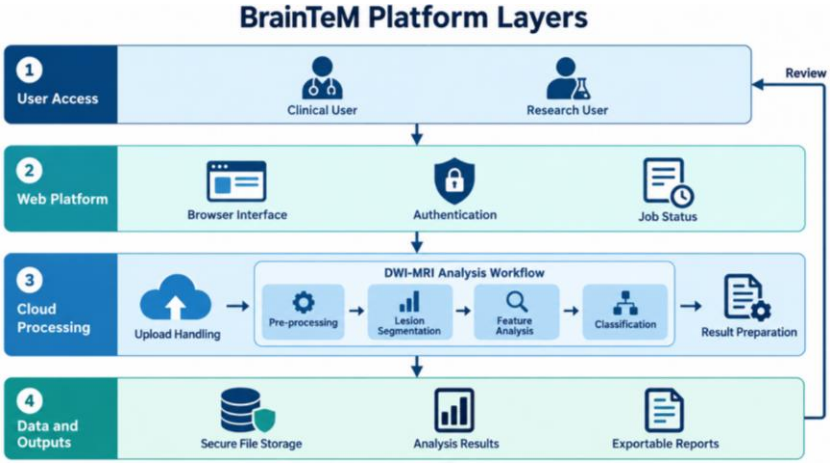


Figure 2. High-level view of the BrainTeM platform layers. The diagram is conceptual and omits deployment-specific configuration.

Table 1. Summary of the analysis workflow stages.

Stage	Input / action	Output
1.Submission	User selects a DICOM file or ZIP package in the browser.	Validated upload and processing request.
2. API handling	Flask service accepts the request and identifies image content.	Analysis job and source image reference.
3. Image analysis	Preprocessing, FCMAC segmentation, feature extraction, and WKNN prediction.	Segmentation mask, measurements, category, confidence.
4. Result persistence	Generated images and metadata are written to cloud storage and job records.	Retrievable analysis artifacts.
5. Review	Browser displays original, processed, segmented, and overlay views with metrics.	Structured output for research review.

3.2 Evaluation and reporting of medical AI

The project inventory lists 166 cases from three sources: 67 acute cases from ISLES 2015, approximately 53 cases from Malaysian hospitals (20 acute and 33 chronic), and 46 acute cases contributed through PadiMedical/Lönge Medikal. These data sources and the related segmentation and classification work were described in our previous study [10]. The authors should specify which data and model outputs are reused here, and which evaluations were conducted specifically for the cloud platform. Neuroradiology input was used to delineate the ground-truth lesion regions, according to the project records. The current workflow processes select DWI slices; it is not a full volumetric analysis of every case. Before submission, the investigators should reconcile the cohort definitions and the number of slices used for model development against the final dataset manifest.

These collections differ in how they were acquired and curated. ISLES is a public challenge benchmark [3], whereas the hospital and partner datasets were assembled for this project. The supplied project record did not include patient demographics, scanner distributions, ethics identifiers, or consent details. The investigators should add the relevant approvals, data-use permissions, and de-identification procedures before submitting the manuscript.

3.3 Preprocessing and lesion segmentation

DWI images undergo intensity normalization, background removal, and gamma enhancement. Normalization reduces differences in intensity scale, background removal suppresses non-brain pixels, and gamma adjustment changes image contrast for subsequent analysis. The processing parameters depend on the configuration and should be recorded alongside the versioned code used in any reproducibility study.

FCMAC combines fuzzy c-means clustering with active-contour refinement to identify a candidate lesion region. FCM assigns soft membership values to pixels, helping identify regions with similar intensity; the active contour then refines the candidate boundary. The resulting binary mask is used for visualization and feature calculation. In this prototype, it represents an algorithm-generated candidate region and requires confirmation by a qualified reader. The overall sequence, from input through image preparation and lesion analysis to measurements and result review, is summarized in Figure 3. The diagram presents the workflow at a high level and omits implementation-specific parameters.

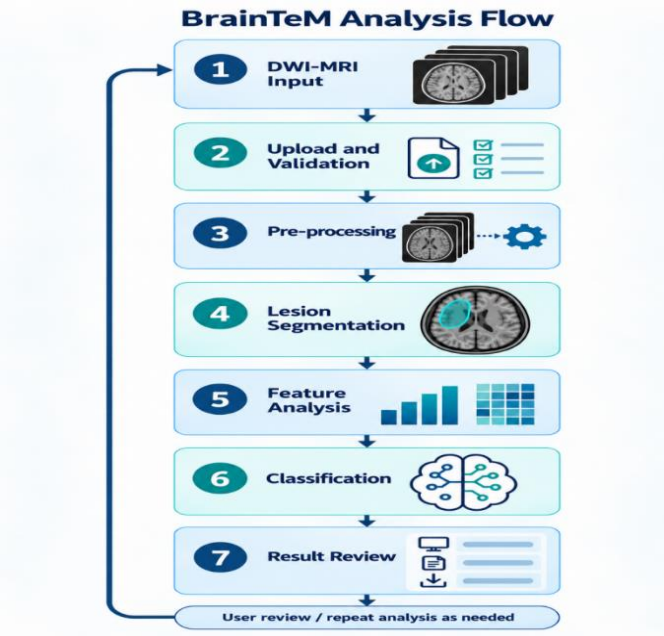


Figure 3. Abstract analysis flows from image input to result review.

3.4 Feature extraction and classification

The segmented region is described by 13 measurements covering its size and boundary geometry, intensity distribution, and selected histogram or gradient characteristics. These values form the input to a weighted K-nearest-neighbour

classifier, which returns the class encoded in the trained model and a confidence value. This value is a model output, not a calibrated probability of disease.

Table 2. The features that are used in project pipeline.

Feature group	Features
Region geometry	Area; perimeter; compactness.
Intensity summary	Mean; maximum normalized intensity; standard deviation; median; mode within ROI; mean boundary intensity.
Distribution and gradients	Skewness; kurtosis; mean gradient magnitude (MGR); variance of gradient response (VGR).

BrainTeM retains the 13-feature set used in the established project pipeline. To reproduce it precisely, the implementation record should define compactness, MGR, and VGR, and specify the normalization conventions and units for each feature.

3.5 Cloud implementation and result presentation

The web service provides routes for the landing page, analysis submission, DICOM preview, model information, and retrieval or deletion of uploaded analysis records. The backend processes DICOM files and ZIP packages, then makes the source, preprocessed, segmented, and overlay images, along with numeric outputs, available to the browser through cloud-backed storage. The project uses Cloud Run to host the containerized Flask service and Firebase components for storage and application data [35–36]. Users can access the workflow remotely without installing the analysis software on their own machines.

For review, the interface can show the source, processed, segmented, and overlay images, together with the predicted label, confidence value, extracted features, and available segmentation metrics. Depending on the analysis record, these metrics may include Dice coefficient, true-positive rate, false-negative rate, precision, and accuracy. Their appearance in the interface does not mean that every metric has been independently measured for every uploaded case.

3.6 Evaluation approach

We considered functional evidence separately from model evidence. Functional evidence concerns whether the deployed workflow accepts supported inputs, runs the analysis stages, stores the outputs, and displays the results. The saved model information lists 93.10% test accuracy, 93.86% cross-validation accuracy, and 65.08% balanced accuracy. These measures are not interchangeable: with imbalanced classes, overall accuracy can remain high even when average class recall is lower [37].

The records available for this paper did not specify the test-set size, patient-level split, fold construction, class counts, confusion matrix, confidence intervals, or external test protocol. We therefore report the scores only as summaries supplied by the deployed model. This study did not include an independent user study, runtime benchmark, prospective clinical assessment, or comparison with radiologist decisions.

3.7 Functional scope of the prototype

The prototype supports a small set of user tasks: upload a supported image package, start processing, check its status, inspect the image views, and review the resulting

measurements and class. We have not measured usability formally or tested every DICOM encoding and ZIP structure. Before the system is used with new clinical collections, input validation, series selection, access control, and storage behavior should be assessed against a documented acceptance set.

A reproducible release should record the software version, model artifact, preprocessing settings, service configuration, and output schema. These details are needed for controlled evaluation, although this platform paper reports the deployment only at a high level.

Results and Discussion

System design

BrainTeM links image submission and result review to a multistage analysis pipeline. A user can upload a DICOM image or ZIP package, start an analysis job, and inspect the resulting image views and measurements in a browser. The engineering contribution is the delivery of preprocessing, candidate lesion segmentation, feature-based classification, and visualization through a single remote workflow. Because results are stored separately from the web service, they remain available after processing finishes.

From the user's perspective, the workflow has four phases: submission, processing, result storage, and review. Keeping these phases distinct makes it easier to identify whether an issue occurred before analysis, during processing, or while displaying the result. It also offers a starting point for future logging and service monitoring; no latency or uptime measurements are reported here.

The system bridges an offline analysis pipeline and a browser-accessible application. Showing intermediate images gives reviewers a way to judge whether the candidate region appears plausible; a label or score alone would not show where the model focused. Overlays can also appear more certain than they are, so the interface should identify the output as a candidate result for research review.

4.2 Model summary and interpretation

The model that has been used in the pipeline record reports 93.10% accuracy and 65.08% balanced accuracy as stated in Table 3. This difference cautions against relying on overall accuracy alone, particularly if some classes are under-represented. Without the class distribution and confusion matrix, we cannot determine which class contributes most to the difference or independently calculate sensitivity and specificity. These values should not be described as “93% clinically accurate” or compared directly with studies using patient-level external tests or lesion-overlap measures.

Segmentation also needs its own quantitative evaluation. A suitable study would compare each case with the expert reference mask and report the Dice similarity coefficient alongside measures such as lesion-wise sensitivity, precision, and volume difference. Case-level distributions are more informative than a pooled average alone. The study should also document how annotations were made, inter-reader variation, and how small lesions were handled. Classification accuracy cannot stand in for segmentation validation.

Table 3. Model summary

Reported model measure	Value	Interpretation
Test accuracy	93.10%	Overall proportion correctly classified in the recorded evaluation; test-set size and split details unavailable.
Cross-validation accuracy	93.86%	Reported mean or summary accuracy; fold-level values and validation protocol unavailable.
Balanced accuracy	65.08%	Class-balanced summary; materially lower than overall accuracy and consistent with class-wise performance needing scrutiny.

4.3 Result presentation and review boundary

The result bundle keeps the source image, prepared image, candidate mask, overlay, and model summary together for inspection (Figure 4). This lets a reader see where the algorithm marked a region before considering its category or confidence value. BrainTeM does not produce a radiology report; its confidence value is not a calibrated probability of disease. Interpretation in the full clinical context remains the responsibility of a qualified reader.

Figure 4 groups the output components conceptually. It is not a replica of the live interface, does not show code-level processing settings, and does not imply that every view is available for every upload. Detailed model analysis and clinical evaluation are outside the scope of this platform paper and are better addressed in separate studies.



Figure 4. Schematic of result components presented for review.

4.4 Operational considerations

Research data may arrive as individual DICOM images or as ZIP packages, so both input types are useful. Reliable processing also requires checks for malformed files, unsupported transfer syntaxes, missing pixel data, mixed studies, and unexpected series contents. The service should identify which slice or series it selected and show processing status clearly, especially when a package contains several possible images. An audit record should connect the source file with the software and model versions, processing time, outputs, and any error message.

Cloud hosting makes remote access and centralized updates possible, but it also raises governance questions. Medical images can retain identifying metadata even after pixel-level de-identification. Before clinical data are used, the project should verify access controls, retention periods, encryption, audit trails, backups, institutional approvals, and relevant Malaysian privacy and health-data requirements. This paper does not claim regulatory clearance or production-grade security certification.

4.5 Limitations and next validation steps

The evidence is limited by the size and heterogeneity of the project collection, the use of selected slices, incomplete details about the classifier split, and the absence of an independent external test. Slices from the same patient may be correlated; a random slice-level split could place a patient's images in both training and test sets and inflate the reported scores. Future evaluation should split by patient and, where possible, by institution or acquisition source. A locked evaluation script should report cohort flow, class counts, preprocessing, model parameters, fold design, the confusion matrix, class-wise sensitivity and specificity, balanced accuracy, and confidence intervals.

The web workflow has not undergone formal usability testing or prospective assessment in a clinical setting. Future studies should include neuroradiologists and intended users, define the tool's role, record task completion and error rates, and examine how users interpret the overlay and confidence value. Any 3D extension should be evaluated separately using volumetric inputs and patient-level ground truth; it should not be presented as part of the current selected-slice prototypede.

Conclusion

This paper describes BrainTeM, a cloud-hosted prototype that connects DWI upload with image preparation, FCMAC candidate segmentation, 13-feature extraction, weighted k-nearest-neighbour classification, and browser-based result review. The application brings intermediate images and model outputs together in a remote workflow. The deployed model record lists 93.10% test accuracy, 93.86% cross-validation accuracy, and 65.08% balanced accuracy. The gap between overall and balanced accuracy underlines the need for class-wise evaluation. Because the patient-level split, confusion data, and external validation were not available, these figures remain preliminary system summaries and do not establish clinical effectiveness. BrainTeM is intended for research support and structured review. Independent multi-site validation, transparent reporting, usability testing, and governance review are needed before clinical deployment.

Acknowledgement
The authors would like to thank Universiti Teknikal Malaysia Melaka (UTeM), Faculty of Electrical Technology and Engineering (FTKE), Faculty of Electronic Technology and Engineering (FTKEK), Grant IMAP/2024/FTKEK/PADI MEDICAL/IM0014, PadiMedical, and Lönge Medikal for their collaboration and support. The authors also acknowledge the clinical experts who contributed to the review and annotation of the DWI-MRI images. This work was supported by the IMaP grant [confirm the official grant title and number]. The authors are grateful to all project collaborators who contributed to the development of the BrainTeM platform.

References

[1] Schellinger, P.D., et al. Evidence-based guideline: The role of diffusion and perfusion MRI for the diagnosis of acute ischemic stroke: Report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. Neurology 75(2), 177–185 (2010). https://doi.org/10.1212/WNL.0b013e3181e7c9dd.

[2] van Everdingen, K.J., van der Grond, J., Kappelle, L.J., Ramos, L.M.P., Mali, W.P.T.M. Diffusion-weighted magnetic resonance imaging in acute ischemic stroke. Stroke 29(9), 1783-1790 (1998). https://doi.org/10.1161/01.STR.29.9.1783.

[3] Maier, O., et al. ISLES 2015 - A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. Medical Image Analysis 35, 250-269 (2017). https://doi.org/10.1016/j.media.2016.07.009.

[4] GBD 2021 Stroke Risk Factor Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990-2021: A systematic analysis for the Global Burden of Disease Study 2021. The Lancet Neurology 23(10), 973-1003 (2024). https://doi.org/10.1016/S1474-4422(24)00369-7.

[5] Feigin, V.L., Brainin, M., Norrving, B., Martins, S.O., Pandian, J., Lindsay, P., Grupper, M., Rautalin, I. World Stroke Organization: Global Stroke Fact Sheet 2025. International Journal of Stroke 20(2), 132–144 (2025). https://doi.org/10.1177/17474930241308142.

[6] Noor, N.S.M., Saad, N.M., Abdullah, A.R., Ali, N.M. Automated segmentation and classification technique for brain stroke. International Journal of Electrical and Computer Engineering 9(3), 1832-1841 (2019). https://doi.org/10.11591/ijece.v9i3.pp1832-1841.

[7] Liew, S.L., et al. A large, curated, open-source stroke neuroimaging dataset to improve lesion segmentation algorithms. Scientific Data 9, 320 (2022). https://doi.org/10.1038/s41597-022-01401-7.

[8] Roberts, M., et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. Nature Machine Intelligence 3, 199-217 (2021). https://doi.org/10.1038/s42256-021-00307-0.

[9] Azman, I.H., Saad, N.M., Abdullah, A.R., Hamzah, R.A., Samsudin, A.,

Kandaya, S.A.P. Automated CAD System for Early Stroke Diagnosis: Review. International Journal of Advanced Computer Science and Applications 14(8), 77–83 (2023). https://thesai.org/Downloads/Volume14No8/Paper_9-Automated_CAD_System_for_Early_Stroke_Diagnosis.pdf

[10] Azman, I.H., Saad, N.M., Abdullah, A.R., Hamzah, R.A., Muda, A.S., Yamba, F.A. Development of a CAD system for stroke diagnosis using machine learning on DWI-MRI images. International Journal of Advanced Technology and Engineering Exploration 12(129), 1246–1263 (2025). https://doi.org/10.19101/IJATEE.2024.111102264

[11] Hernandez Petzsche, M.R., de la Rosa, E., Hanning, U., et al. ISLES 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset. Scientific Data 9, 762 (2022). https://doi.org/10.1038/s41597-022-01875-5

[12] de la Rosa, E., et al. DeepISLES: a clinically validated ischemic stroke segmentation model from the ISLES'22 challenge. Nature Communications 16, 7357 (2025). https://doi.org/10.1038/s41467-025-62373-x

[13] Pacal, I., Algarni, A., Kunduracioglu, I. Adaptive and efficient deep learning model for automated ischemic stroke lesion segmentation. Biomedical Signal Processing and Control 118, 109658 (2026). https://doi.org/10.1016/j.bspc.2026.109658

[14] Baaklini, M., Valdés Hernández, M.C. Deep learning for MRI-based acute and subacute ischaemic stroke lesion segmentation—a systematic review, meta-analysis, and pilot evaluation of key results. Frontiers in Medical Technology 7, 1491197 (2025). https://doi.org/10.3389/fmedt.2025.1491197.

[15] Bezdek, J.C. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York (1981).

[16] Kass, M., Witkin, A., Terzopoulos, D. Snakes: Active contour models. International Journal of Computer Vision 1, 321-331 (1988). https://doi.org/10.1007/BF00133570.

[17] Cover, T.M., Hart, P.E. Nearest neighbor pattern classification. IEEE Transactions on Information Theory 13(1), 21-27 (1967). https://doi.org/10.1109/TIT.1967.1053964.

[18] National Electrical Manufacturers Association. Digital Imaging and Communications in Medicine (DICOM) Standard, PS3 (2024). Available: https://www.dicomstandard.org/current.

[19] Kelly, C.J., Karthikesalingam, A., Suleyman, M., Corrado, G., King, D. Key challenges for delivering clinical impact with artificial intelligence. BMC Medicine 17, 195 (2019). https://doi.org/10.1186/s12916-019-1426-2.

[20] Collins, G.S., Moons, K.G.M. Reporting of artificial intelligence prediction models. The Lancet 393(10181), 1577-1579 (2019). https://doi.org/10.1016/S0140-6736(19)30037-6.

[21] Langlotz, C.P., et al. A roadmap for foundational research on artificial intelligence in medical imaging: From the 2018 NIH/RSNA/ACR/The Academy Workshop. Radiology 291(3), 781-791 (2019). https://doi.org/10.1148/radiol.2019190613.

[22] Dikici, E., Bigelow, M., Prevedello, L.M., White, R.D., Erdal, B.S. Integrating AI into radiology workflow: Levels of research, production, and feedback maturity. Journal of Medical Imaging 7(1), 016502 (2020). https://doi.org/10.1117/1.JMI.7.1.016502.

[23] World Health Organization. Ethics and Governance of Artificial Intelligence for Health: WHO Guidance. Geneva: World Health Organization (2021). ISBN 978-92-4-002920-0.

[24] National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1 (2023). https://doi.org/10.6028/NIST.AI.100-1.

[25] International Medical Device Regulators Forum. Software as a Medical Device (SaMD): Clinical Evaluation. IMDRF/SaMD WG/N41FINAL:2017 (2017).

[26] U.S. Food and Drug Administration, Health Canada, and Medicines and Healthcare products Regulatory Agency. Good Machine Learning Practice for Medical Device Development: Guiding Principles (2021).

[27] Park, S.H., Han, K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. Radiology 286(3), 800-809 (2018). https://doi.org/10.1148/radiol.2017171920.

[28] Liu, X., et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. Nature Medicine 26, 1364-1374 (2020). https://doi.org/10.1038/s41591-020-1034-x.

[29] Cruz Rivera, S., et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: The SPIRIT-AI extension. Nature Medicine 26, 1351-1363 (2020). https://doi.org/10.1038/s41591-020-1037-7.

[30] Collins, G.S., et al. TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 385, e078378 (2024). https://doi.org/10.1136/bmj-2023-078378.

[31] Vasey, B., et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nature Medicine 28(5), 924-933 (2022). https://doi.org/10.1038/s41591-022-01772-9.

[32] Amann, J., Blasimme, A., Vayena, E., Frey, D., Madai, V.I. Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. BMC Medical Informatics and Decision Making 20, 310 (2020). https://doi.org/10.1186/s12911-020-01332-6.

[33] Ghassemi, M., Oakden-Rayner, L., Beam, A.L. The false hope of current approaches to explainable artificial intelligence in health care. The Lancet Digital Health 3(11), e745-e750 (2021). https://doi.org/10.1016/S2589-7500(21)00208-9.

[34] Topol, E.J. High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine 25, 44-56 (2019). https://doi.org/10.1038/s41591-018-0300-7.

[35] Google Cloud. Cloud Run documentation. Available:

https://cloud.google.com/run/docs (accessed 19 October 2026).

[36] Google. Firebase documentation. Available: https://firebase.google.com/docs (accessed 19 October 2026).

[37] Saito, T., Rehmsmeier, M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLOS ONE 10(3), e0118432 (2015). https://doi.org/10.1371/journal.pone.0118432