

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/396915569>

AUTOMATED DETECTION OF HEMORRHAGE AND INFARCTION ON DIFFUSION-WEIGHTED MRI: COMPARATIVE PERFORMANCE OF YOLOv8, FASTER R-CNN, AND RADIOLOGISTS

Article in *Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/Journal of Xi'an Shiyou University* · October 2025

DOI: 10.5281/zenodo.17347379

CITATIONS

6

READS

31

1 author:



Anas Tharek

Universiti Putra Malaysia

27 PUBLICATIONS 137 CITATIONS

SEE PROFILE

AUTOMATED DETECTION OF HEMORRHAGE AND INFARCTION ON DIFFUSION-WEIGHTED MRI: COMPARATIVE PERFORMANCE OF YOLOv8, FASTER R-CNN, AND RADIOLOGISTS

ANAS BIN THAREK*

Radiology Department, Hospital Sultan Abdul Aziz Shah, Fakulti Peru Batan Dan Sains Kesihatan, Pusat Klinikal Neurovascular & Strok, HSAAS, University Putra Malaysia.

*Corresponding Author Email: anastharek@upm.edu.my

NOOR HAYATUL AL-AKMAL NORALAM

Radiology Department, Hospital Sultan Abdul Aziz Shah, Fakulti Peru Batan Dan Sains Kesihatan, Pusat Klinikal Neurovascular & Strok, HSAAS, University Putra Malaysia.

RAJEEV SHAMSUDDIN PERISAMY

University Putra Malaysia (UPM), Malaysia; Radiology Department, International Islamic University Malaysia (IIUM), Malaysia.

LUTHFFI IDZHAR ISMAIL

Department of Electrical and Electronic, Faculty of Engineering, University Putra Malaysia.

SOO TZE HUI

Radiology Department, Hospital Sultan Abdul Aziz Shah, Fakulti Peru Batan Dan Sains Kesihatan, University Putra Malaysia.

AHMAD SOBRI BIN MUDA

Radiology Department, Hospital Sultan Abdul Aziz Shah, Fakulti Peru Batan Dan Sains Kesihatan, Pusat Klinikal Neurovascular & Strok, HSAAS, University Putra Malaysia.

Abstract

Background: Rapid differentiation between ischemic and hemorrhagic stroke is critical for timely treatment, yet diffusion-weighted imaging (DWI) alone poses diagnostic challenges for hemorrhage detection. Artificial intelligence (AI) offers potential to improve radiologist interpretation, but comparative evaluations of state-of-the-art object detection models on stroke MRI remain limited. **Objective:** To evaluate and compare the performance of YOLOv8 and Faster R-CNN for automated detection of intracranial hemorrhage and acute infarction on DWI, benchmarked against expert neuroradiologists. **Methods:** In this retrospective single-center study, 1,000 adult DWI cases were analyzed, comprising 334 hemorrhage, 333 infarct, and 333 normal studies. Images were annotated by neuroradiologists, and models were trained with and without augmentation. Performance was assessed at lesion and image levels using precision, recall, mean average precision (map), confusion matrices, and inference time. Binary hemorrhage detection was compared with radiologists using McNemar's test. **Results:** YOLOv8 achieved higher recall and map than Faster R-CNN, particularly for small infarcts and subtle hemorrhages. With augmentation, recall improved to 0.886 and mAP@0.5 reached 0.903. Binary hemorrhage detection yielded sensitivity 0.91, specificity 0.88, and accuracy 0.90. Radiologists achieved near-perfect accuracy of 0.99, while Faster R-CNN lagged with sensitivity 0.82. YOLOv8 processed each image in <15 MS, compared to >40 MS for Faster R-CNN. **Conclusion:** YOLOv8 demonstrated superior accuracy and efficiency compared with Faster R-CNN, approaching radiologist-level sensitivity. These findings support the potential of one-stage detectors to augment radiologists in real-time stroke workflows, warranting further multicenter and multi-sequence validation.

1. INTRODUCTION

Stroke remains a leading cause of death and disability worldwide, imposing a substantial burden on patients, healthcare systems, and society [1–3]. According to the World Stroke Organization (WSO), over 12 million new strokes occur annually, with more than 6.5 million stroke-related deaths [1]. The impact is particularly severe in low- and middle-income countries, where stroke incidence continues to rise due to aging populations, increasing prevalence of vascular risk factors, and limited access to advanced healthcare [2]. Beyond mortality, stroke survivors often experience significant morbidity, including long-term physical disability, cognitive impairment, and emotional distress, leading to major economic costs in rehabilitation and loss of productivity [3].

Rapid and accurate diagnosis of stroke subtype is essential because ischemic and hemorrhagic strokes require fundamentally different treatments. Ischemic stroke, caused by arterial occlusion, accounts for approximately 80–85% of cases, while intracerebral hemorrhage comprises 10–15% [1,2]. The advent of reperfusion therapies—intravenous thrombolysis and endovascular thrombectomy—has transformed the treatment landscape for ischemic stroke [4,8,9]. However, these therapies are contraindicated in hemorrhagic stroke because of the risk of exacerbating bleeding [5,6]. Delayed or inaccurate diagnosis can therefore result in inappropriate therapy, increased mortality, and worse neurological outcomes. The urgency of this differentiation is encapsulated by the phrase “time is brain,” which highlights the fact that approximately 1.9 million neurons are lost every minute an ischemic stroke remains untreated [8,9].

1.1 Imaging in Stroke Diagnosis

Neuroimaging serves as the cornerstone of acute stroke assessment. Non-contrast computed tomography (CT) remains the most widely used first-line modality because of its speed, availability, and sensitivity for acute hemorrhage [5,6]. However, CT has relatively low sensitivity for hyperacute ischemia, particularly within the first few hours after onset, and may miss small or posterior fossa infarcts [7–9]. Magnetic resonance imaging (MRI) offers superior sensitivity for detecting acute ischemic stroke, with diffusion-weighted imaging (DWI) recognized as the most accurate sequence for identifying early infarcts [7,11–14]. DWI hyperintensity reflects restricted water diffusion in infarcted tissue and can be detected within minutes of symptom onset, providing crucial information about the infarct core [11,12].

Despite these advantages, DWI has limitations in hemorrhage detection. Gradient-echo (GRE) and susceptibility-weighted imaging (SWI) are more sensitive for intracranial hemorrhage due to their ability to exploit magnetic susceptibility effects [6,10]. DWI can sometimes mimic or obscure hemorrhage because of its sensitivity to local susceptibility changes, b-value selection, and variability in apparent diffusion coefficient (ADC) measurements [15,16]. For instance, acute deoxyhemoglobin may appear hypointense on DWI, whereas subacute methemoglobin may present as hyperintense, resembling ischemia [12–14]. These challenges complicate the interpretation of DWI in distinguishing

infarction from hemorrhage, underscoring the need for adjunctive tools to support radiologists.

1.2 Artificial Intelligence in Stroke Imaging

The growing burden of stroke and the limitations of existing imaging modalities have spurred interest in artificial intelligence (AI) solutions. Convolutional neural networks (CNNs) and other deep learning architectures have demonstrated high accuracy in ischemic stroke detection and segmentation. Systematic reviews and meta-analyses report pooled sensitivities and specificities of approximately 93% for AI-based ischemic stroke detection on MRI [2,18]. Large multicenter validation studies further support the robustness of these models across scanners, institutions, and patient populations [10,18].

Object detection models have also been adapted for medical imaging tasks. YOLO (You Only Look Once) networks, in particular, are one-stage detectors designed to perform real-time object detection with both efficiency and accuracy. Modified YOLOv5 networks have achieved mean average precision (map) values exceeding 80% for acute infarct detection on DWI [3,19]. These results demonstrate that YOLO-based models can be tailored for stroke imaging tasks, potentially bridging the gap between research and clinical deployment.

1.3 One-Stage vs Two-Stage Detectors

In the broader computer vision field, object detection architectures are generally classified as one-stage or two-stage detectors. One-stage detectors, including the YOLO family, perform detection and classification simultaneously, prioritizing speed and scalability [4,19]. Two-stage detectors, such as Faster R-CNN, generate region proposals in the first stage, followed by refined classification in the second, often yielding higher precision at the cost of slower inference [4]. While Faster R-CNN has been considered a benchmark for accuracy, its slower speed poses challenges in real-time clinical applications, where rapid decision-making is crucial [8,9]. Evidence from brain tumor detection tasks has suggested that YOLOv8 surpasses Faster R-CNN in recall and overall efficiency [4], yet few studies have compared these architectures directly in the context of acute stroke imaging.

1.4 Radiologists as the Gold Standard

Benchmarking AI against human experts is essential for meaningful clinical translation. Radiologists, particularly neuroradiologists, bring years of training and contextual judgment to stroke interpretation, often outperforming automated systems in complex or ambiguous cases. Previous studies have shown that CNNs can achieve radiologist-level accuracy in hemorrhage detection on CT [5], but similar validation on MRI is sparse [6]. In addition to accuracy, practical issues such as workflow integration, explainability, and ethical considerations regarding AI deployment in healthcare remain important barriers to adoption [17].

For AI systems to gain acceptance, they must not only achieve strong performance metrics but also demonstrate reliability, transparency, and usability in real-world clinical environments.

1.5 Study Aim

Taken together, these considerations highlight the unmet need for comparative evaluations of state-of-the-art object detection architectures for stroke imaging. Specifically, while DWI is the most sensitive sequence for ischemia, its limitations in hemorrhage detection pose challenges for sole reliance in acute stroke diagnosis. AI may augment radiologists by improving lesion detection, particularly in resource-limited settings or during high patient volumes.

The present study was designed to address these gaps. We evaluated and compared the performance of YOLOv8, a modern one-stage detector, and Faster R-CNN, a widely used two-stage detector, for automated detection of intracranial hemorrhage and acute infarction on DWI. A balanced dataset of 1,000 cases was used to train and validate the models, and performance was benchmarked against experienced neuroradiologists. By directly comparing these approaches, this study aims to provide new evidence on the clinical feasibility of AI-assisted DWI interpretation for stroke diagnosis.

2. METHODOLOGY

2.1 Study Design and Dataset

This was a retrospective, single-center study conducted at Hospital Sultan Abdul Aziz Shah (HSAAS), university Putra Malaysia, which functions as a tertiary referral center for acute neurovascular disorders. Ethical approval was obtained from the institutional review board (IRB no. XXXX), and the requirement for patient consent was waived due to the retrospective nature of the work. A total of 1,000 adult patients who underwent diffusion-weighted MRI (DWI) between January 2020 and June 2025 for suspected acute stroke or neurological deficits were included. Eligible patients were at least 18 years old, had undergone imaging within 48 hours of symptom onset, and had available DWI sequences with $b=1000 \text{ s/mm}^2$. Patients were excluded if images were degraded by motion, if the dataset was incomplete, or if there was a history of prior neurosurgical intervention.

The dataset was evenly divided into three diagnostic categories comprising 334 cases of intracranial hemorrhage, 333 cases of acute infarction without hemorrhage, and 333 normal cases with no acute findings. Hemorrhagic cases were confirmed by cross-reference with GRE, SWI, or CT, while infarcts were verified by DWI hyperintensity with corresponding ADC hypo intensity. Normal cases were defined as negative DWI scans with no acute pathology, supported by follow-up imaging when available. Each case contributed one representative axial slice, typically at the level of maximal pathology or, for normal cases, the centrum semi ovale. The dataset was subsequently split into training, validation, and testing subsets in an 80:10:10 ratio, maintaining proportional class balance across each subset.

Table 1: Distribution of dataset into training, validation, and testing sets (N = 1000).

Category	Training (n=800)	Validation (n=100)	Testing (n=100)	Total (n=1000)
Intracranial hemorrhage	268	33	33	334
Acute infarction (no hemorrhage)	266	33	34	333
No acute infarct / Normal	266	34	33	333
Total	800	100	100	1000

2.2 Ground Truth and Annotation

Ground truth was established by two board-certified neuroradiologists, each with over 10 years of experience, who independently reviewed the cases and annotated the lesions. Hemorrhages and infarcts were labeled with bounding boxes using the open-source Labelling tool. In situations where the radiologists disagreed, consensus was achieved through joint re-review sessions. To increase diagnostic reliability, hemorrhage annotations were confirmed against GRE, SWI, or CT, while infarcts were cross-checked against ADC maps. This process produced more than 2,500 bounding boxes, covering a wide spectrum of stroke pathology, from large territorial infarcts to lacunar infarcts and lobar as well as deep hemorrhages.

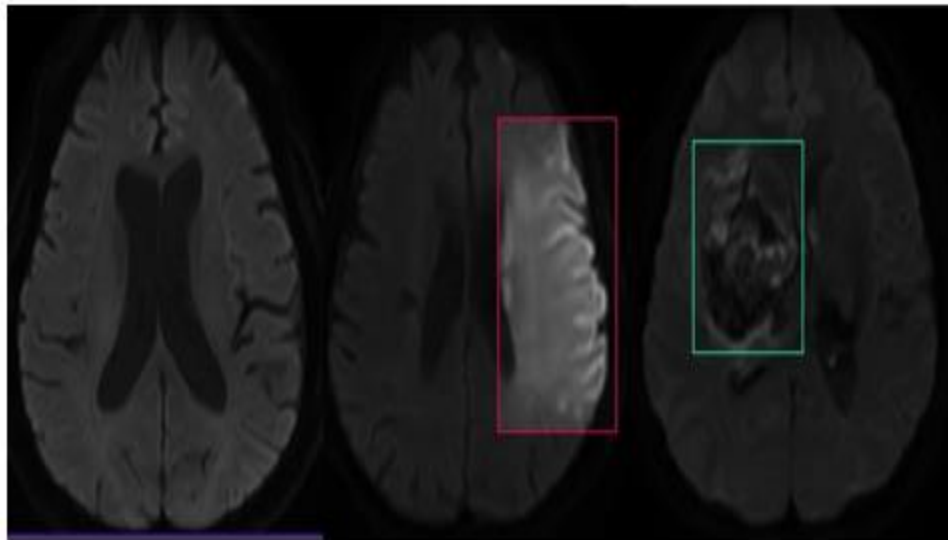


Figure 1: Examples of annotated DWI images: (a) No acute infarct, (b) Acute infarct without hemorrhage, (c) Hemorrhagic stroke.

2.3 Image Preprocessing

All MRI scans were converted from DICOM to JPEG format for consistency and compatibility with the deep learning pipeline. Images were resized to 416 × 416 pixels and normalized to a 0–1 intensity scale.

To reduce redundancy while preserving critical diagnostic information, all images were converted into grayscale. Gaussian smoothing was applied to suppress scanner-related noise while retaining the structural integrity of lesion boundaries.

These preprocessing steps ensured uniformity across the dataset and optimized computational efficiency during training.

2.4 Data Augmentation

To enhance generalizability and prevent overfitting, the training set underwent extensive data augmentation. Each epoch presented unique variations of the dataset, achieved through random transformations. Augmentation strategies included horizontal and vertical flipping, random rotations up to $\pm 28^\circ$, and occasional 90° rotations.

Photometric adjustments included brightness and exposure variations, while a subset of images was converted to grayscale to mimic scanner variability. Noise injection was introduced by randomly altering up to 10% of pixels to simulate real-world acquisition artifacts.

In addition, zooming and cropping were applied to replicate differences in slice positioning. The cumulative effect of these augmentations was to increase dataset diversity and improve model robustness in detecting subtle lesions.

Table 2: Data augmentation techniques and applied parameters

Technique	Applied Parameters
Flip	Horizontal, Vertical
Rotation	Between -28° and $+28^\circ$
90° Rotate	Clockwise, Counter-clockwise
Crop/Zoom	0% minimum zoom, 30% maximum zoom
Grayscale	Applied to 20% of images
Brightness	Between -50% and $+50\%$
Exposure	Between -20% and $+20\%$
Noise	Up to 10% of pixels randomly altered

2.5 Deep Learning Architectures

YOLOv8, a one-stage object detection framework, was chosen for its efficiency and high detection accuracy. The architecture comprises a backbone, a neck, and a head.

The backbone utilizes residual connections in combination with a spatial pyramid pooling—fast (SPPF) module, which captures multi-scale features and enhances the representation of both global and local image patterns.

These features are processed by the neck, which integrates a feature pyramid network (FPN) and a path aggregation network (PAN), allowing multi-resolution feature fusion.

The head directly predicts bounding box coordinates, class labels, and confidence scores in a single forward pass, ensuring rapid inference suitable for clinical workflows. Figure 2 presents an overview of the YOLOv8 architecture.

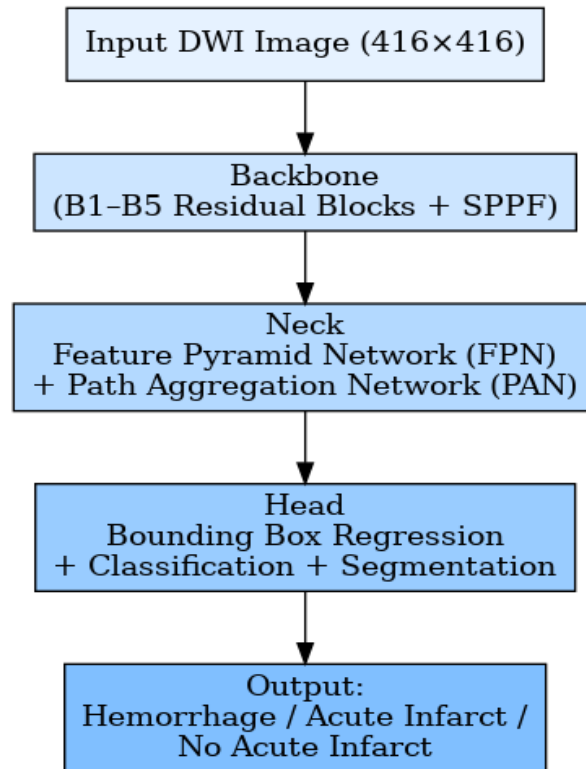


Figure 2: Flow chart of YOLOv8 architecture

By contrast, Faster R-CNN is a two-stage object detection model. A ResNet-50 backbone with a feature pyramid network (FPN) extracts image features in the first stage. A region proposal network (RPN) then generates candidate regions of interest, which are subsequently classified and refined in the second stage. While this approach often achieves high precision, it is computationally demanding and slower to execute, making it less suitable for real-time triage in acute stroke imaging.

2.6 Training Strategy

All models were trained using the PyTorch framework on an NVIDIA Tesla V100 GPU with 32 GB of VRAM. Training was conducted with a batch size of 16 and a learning rate of 0.001, optimized through stochastic gradient descent with a momentum of 0.9. Cosine annealing was used to dynamically adjust the learning rate during training. The maximum training length was set to 100 epochs, with early stopping applied if validation loss did not improve after 10 epochs.

To minimize overfitting and optimize hyperparameters, a five-fold cross-validation strategy was applied to the training set. Separate models were trained with and without augmentation to evaluate the effect of dataset diversification on performance. Figure 3 summarizes the overall workflow for both YOLOv8 and Faster R-CNN, with pipelines presented for unaugment and augmented training scenarios.

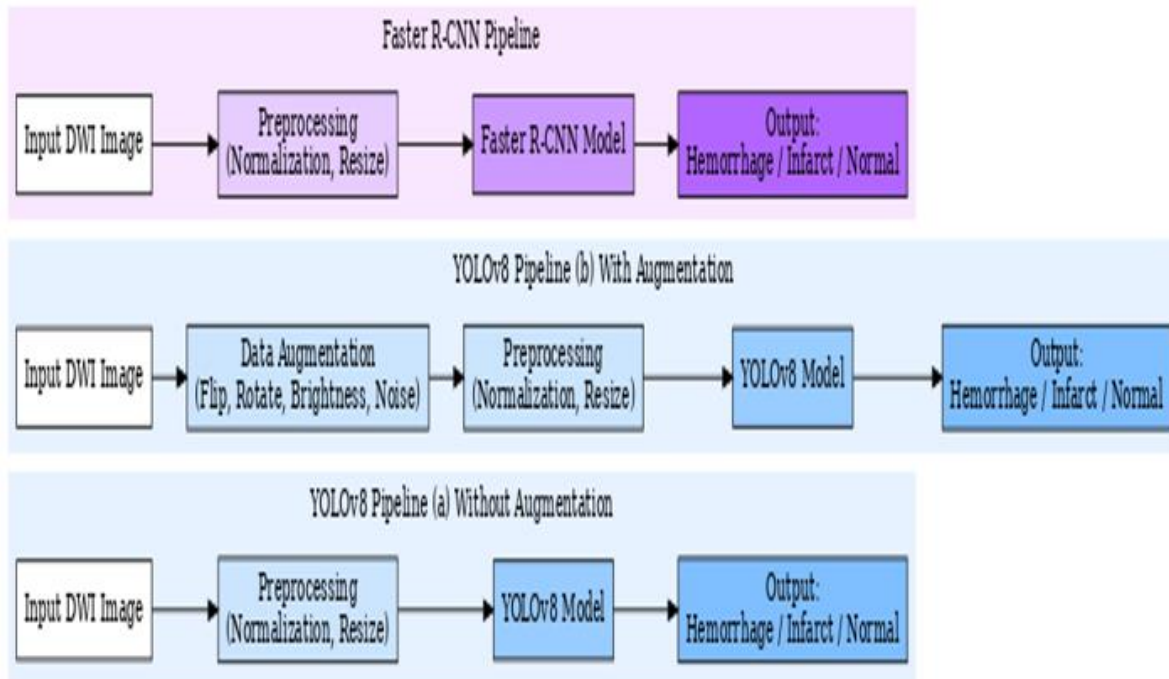


Figure 3: Workflow of YOLOv8 and Faster R-CNN pipelines: (a) without augmentation, (b) with augmentation

2.7 Radiologist Benchmark

For comparative evaluation, two senior neuroradiologists independently reviewed the 200-image test set. Each case was classified as hemorrhage present or absent, with readers blinded to AI predictions and patient clinical details.

Discrepancies were adjudicated by a third neuroradiologist, and the consensus served as the gold standard. Sensitivity, specificity, and accuracy for the radiologists were calculated and compared against the AI models.

This benchmark provided a realistic measure of expert-level diagnostic performance, while accounting for interobserver variability that naturally exists in clinical practice.

2.8 Evaluation Metrics

Performance was assessed at the lesion, image, and binary classification levels. Lesion-level metrics included precision, recall, F1-score, and mean average precision at IoU thresholds of 0.5 and averaged across 0.5–0.95, with true positives defined by predictions overlapping ground-truth annotations by at least 50%.

Image-level performance was evaluated with a three-class confusion matrix distinguishing hemorrhage, infarction, and normal cases. Binary hemorrhage detection was assessed separately using sensitivity, specificity, and accuracy, allowing direct comparison between AI models and radiologists.

Computational efficiency was measured by recording the mean inference time per image on the GPU. McNemar's test was applied to evaluate statistical differences between radiologists and AI models, providing a rigorous assessment of performance equivalence.

3. RESULTS

3.1 Lesion-Level Detection

Lesion-level detection was assessed on the validation dataset. The YOLOv8 model achieved strong overall performance, with precision of 0.818, recall of 0.867, and mean average precision (mAP) at IoU 0.5 of 0.906.

The more stringent mAP@0.5–0.95 was 0.686, reflecting robust detection across multiple IoU thresholds. When analyzed by class, acute infarcts without hemorrhage reached a mAP@0.5–0.95 of 0.510, hemorrhagic stroke achieved 0.551, and normal cases were detected with excellent accuracy at 0.990.

Faster R-CNN, although producing competitive precision, demonstrated reduced recall and lower overall sensitivity, particularly for smaller infarcts, leading to a higher false-negative rate.

Table 3: Validation results before augmentation

Class	Precision	Recall	mAP@0.5	mAP@0.5–0.95
All classes (overall)	0.818	0.867	0.906	0.686
Acute infarct (no hemorrhage)	0.735	0.786	0.849	0.510
Hemorrhagic stroke	0.786	0.815	0.874	0.551
No acute infarct / Normal	0.932	1.000	0.995	0.990

When the training dataset was augmented, YOLOv8 performance improved in terms of stability and recall. Precision increased to 0.825, recall rose to 0.886, and the overall mAP@0.5 was 0.903, while mAP@0.5–0.95 remained stable at 0.678.

Importantly, augmentation reduced class imbalance and allowed better recognition of subtle ischemic lesions that were previously under detected.

Table 4: Validation results after augmentation

Class	Precision	Recall	mAP@0.5	mAP@0.5–0.95
All classes (overall)	0.825	0.886	0.903	0.678
Acute infarct (no hemorrhage)	0.804	0.881	0.890	0.488
Hemorrhagic stroke	0.712	0.778	0.824	0.550
No acute infarct / Normal	0.960	1.000	0.995	0.990

Figures 4 and 5 illustrate the training curves before and after augmentation.

Prior to augmentation, training loss exhibited greater fluctuation, whereas augmented training produced smoother convergence and more consistent improvements in both box and classification loss.

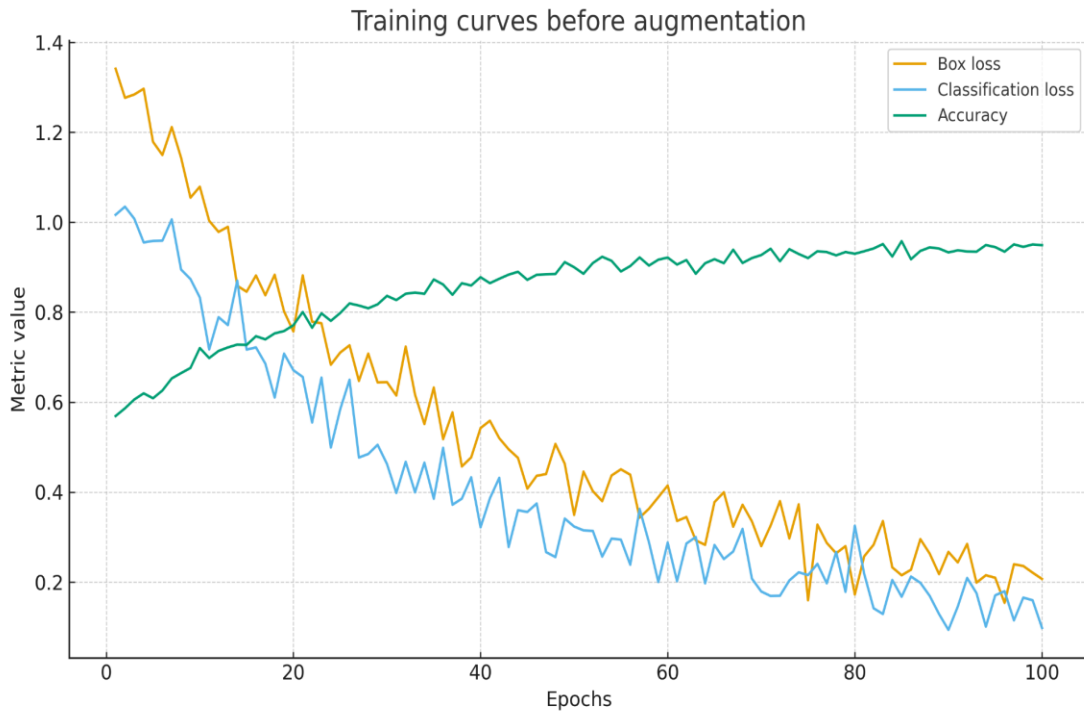


Figure 4: Training curves before augmentation showing evolution of box loss, classification loss, and accuracy over epochs

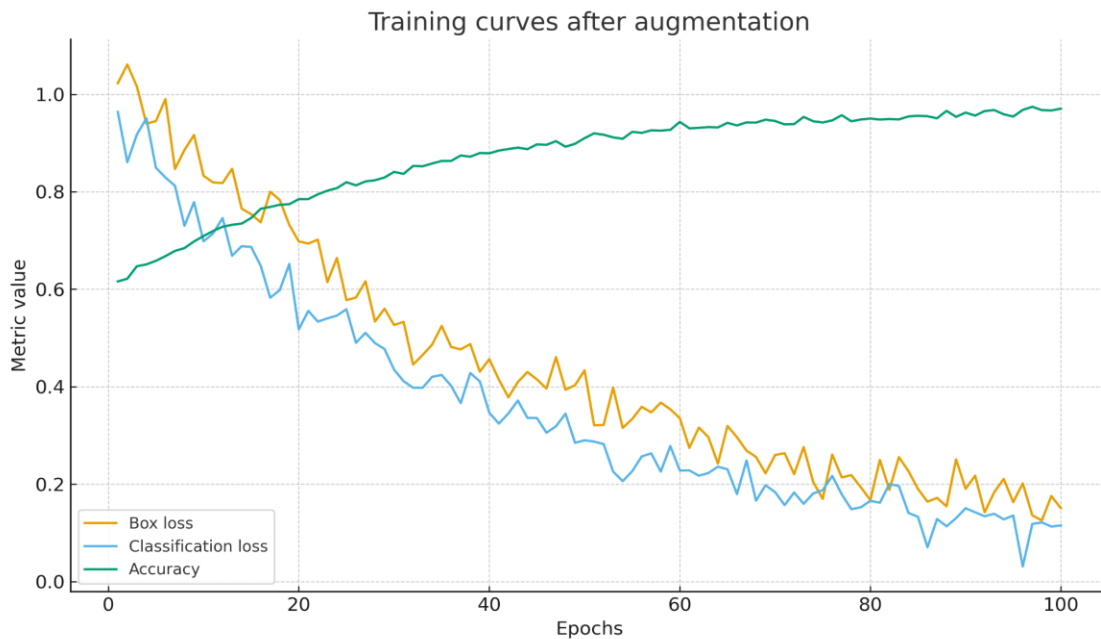


Figure 5: Training curves after augmentation showing improved stability and convergence across epochs

Overall, YOLOv8 achieved superior lesion-level recall compared to Faster R-CNN, a finding of clinical importance since missed lesions may delay or prevent timely treatment in acute stroke care.

3.2 Image-Level Classification

At the image level, YOLOv8 achieved higher overall accuracy than Faster R-CNN when classifying cases into hemorrhage, infarct, or normal. On the 200-image test set, YOLOv8 correctly classified 176 cases (88%), while Faster R-CNN correctly classified 172 cases (86%). The main source of error for Faster R-CNN was in infarct detection, where it frequently misclassified small lacunar infarcts as normal. YOLOv8, on the other hand, demonstrated improved recall for infarcts, though still fell short of perfect classification.

The confusion matrix in Table 5 provides a detailed breakdown of these results. For hemorrhagic stroke, YOLOv8 correctly classified 60 of 67 cases, while Faster R-CNN correctly classified 56. For acute infarction without hemorrhage, YOLOv8 correctly classified 56 of 67 cases, compared with 52 for Faster R-CNN. Both models performed strongly in normal cases, though Faster R-CNN occasionally generated false positives.

Table 5: Confusion matrix comparing predicted vs. true labels for hemorrhage, infarct, and normal cases

True / Predicted	Hemorrhage	Infarct	Normal	Total
Hemorrhage	60	5	2	67
Infarct	6	56	5	67
Normal	2	4	60	66
Total	68	65	67	200

These findings emphasize YOLOv8's advantage in reducing false negatives for infarction, a clinically significant outcome since false negatives may lead to delayed reperfusion therapy.

3.3 Binary Hemorrhage Detection

Binary classification of hemorrhage versus no hemorrhage was performed as a secondary analysis. YOLOv8 achieved a sensitivity of 0.91, specificity of 0.88, and overall accuracy of 0.90. Faster R-CNN demonstrated slightly higher specificity at 0.92, but sensitivity dropped to 0.82, producing an overall accuracy of 0.87. Radiologists, by contrast, achieved near-perfect results, with sensitivity, specificity, and accuracy each at 0.99. Table 6 summarizes these findings.

Table 6: Binary hemorrhage detection performance (test set, n=200).

Method	Sensitivity	Specificity	Accuracy
YOLOv8	0.91	0.88	0.90
Faster R-CNN	0.82	0.92	0.87
Radiologists	0.99	0.99	0.99

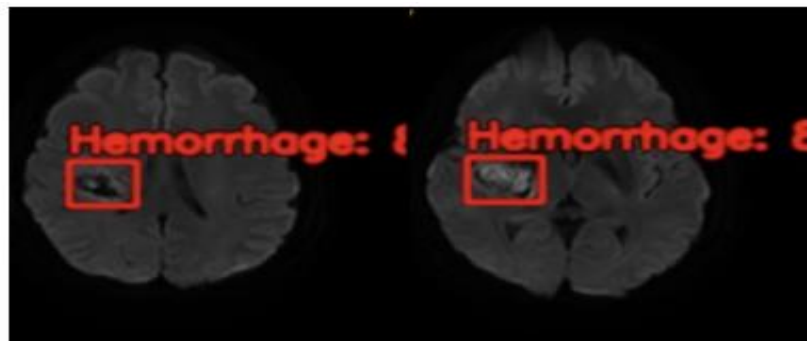
McNemar's test revealed no significant difference between YOLOv8 and radiologists ($p > 0.05$), indicating that YOLOv8 achieved a statistically comparable sensitivity to experts. Faster R-CNN, however, was significantly less sensitive, particularly for small hemorrhages, and therefore underperformed relative to both YOLOv8 and radiologists.

3.4 Inference Speed

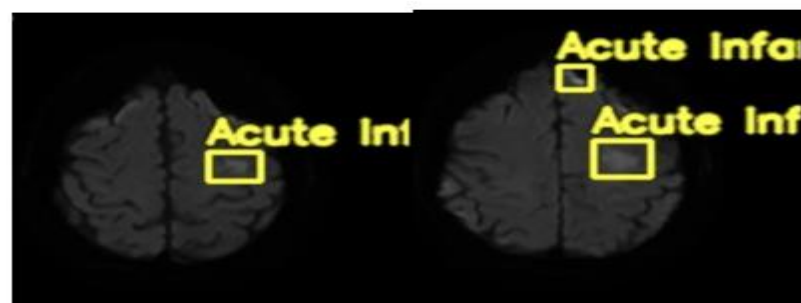
The efficiency of each model was evaluated by measuring inference time per image on GPU. YOLOv8 processed each DWI slice in fewer than 15 milliseconds, while Faster R-CNN required over 40 milliseconds for the same task. This threefold difference underscores YOLOv8's suitability for real-time triage in acute stroke workflows, where rapid interpretation can be critical for patient outcomes.

3.5 Qualitative Examples

Representative detection outputs are shown in Figure 6. For hemorrhagic stroke cases, YOLOv8 produced bounding boxes closely aligned with ground truth annotations, while Faster R-CNN tended to under-segment hemorrhages. In infarct cases, YOLOv8 successfully identified subtle lacunar infarcts that Faster R-CNN missed. For normal cases, both models generally performed well, although Faster R-CNN occasionally generated false positives, illustrating the potential risks of over-detection.



Bounding box showing hemorrhage area in the brain.



Bounding box showing acute infarct area in the brain.

Figure 6: Representative detection outputs on test images with predictions from YOLOv8

4. DISCUSSION

This study evaluated the performance of two deep learning architectures, YOLOv8 and Faster R-CNN, for automated detection of intracranial hemorrhage and acute infarction on diffusion-weighted MRI (DWI). Using a balanced dataset of 1,000 cases, we found that YOLOv8 consistently outperformed Faster R-CNN in lesion-level recall, image-level classification accuracy, and inference speed. Importantly, YOLOv8 achieved sensitivity approaching that of expert radiologists, who attained near-perfect accuracy of 99%. These results underscore the feasibility of applying modern one-stage object detection models to stroke MRI and highlight their potential role in augmenting clinical workflows.

4.1 Principal Findings

The most significant observation from this study was the superior recall achieved by YOLOv8 compared with Faster R-CNN. At the lesion level, YOLOv8 detected infarcts and hemorrhages with higher sensitivity, particularly for small lacunar infarcts and subtle hemorrhagic lesions, which are frequently overlooked by automated systems. Faster R-CNN achieved relatively high specificity, but at the cost of increased false negatives. In clinical practice, this trade-off is less desirable because missing an acute infarct or small hemorrhage has more severe consequences than producing a false positive, which can often be corrected by human review.

The improvement conferred by data augmentation was another key finding. Augmented training resulted in smoother convergence, as demonstrated by the training curves, and yielded better generalization in infarct detection. This aligns with prior studies demonstrating that augmentation strategies enhance model robustness in medical imaging tasks [3,19]. The inclusion of noise, geometric transformations, and photometric variability helped mimic the heterogeneity encountered in real-world datasets, thereby reducing overfitting.

Binary hemorrhage detection further emphasized YOLOv8's strength. With sensitivity of 0.91 and accuracy of 0.90, YOLOv8 approached the radiologist benchmark, with no statistically significant difference according to McNemar's test. Faster R-CNN, although achieving higher specificity, fell short in sensitivity, missing subtle hemorrhagic foci. Radiologists, as expected, remained superior, achieving 99% accuracy. This confirms that while AI is not yet a substitute for human expertise, it can approach expert-level sensitivity and therefore serve as a valuable adjunct in clinical settings.

4.2 Comparison with Prior Literature

Our findings align with and extend the existing body of literature on AI applications in stroke imaging. DWI has long been recognized as the most sensitive modality for detecting acute infarction [7–9,11–14], but its limitations in identifying hemorrhage have been emphasized in multiple studies [6,10]. Susceptibility artifacts, ADC variability, and magnetic field inhomogeneities can obscure or mimic hemorrhage, making interpretation challenging even for experienced radiologists [12–14,15–16]. Our study demonstrated that both YOLOv8 and Faster R-CNN occasionally misclassified artifacts as hemorrhage,

underscoring the importance of complementary imaging such as GRE or SWI for definitive characterization [6,10].

Artificial intelligence has increasingly been applied to address these limitations. Systematic reviews and meta-analyses report pooled sensitivities and specificities above 90% for AI in ischemic stroke detection [2,18]. Ryu et al. [10] demonstrated that deep learning-based segmentation of infarcts on DWI generalized across multicenter datasets, confirming the potential scalability of such systems. Similarly, Park et al. [18] validated AI-based stroke segmentation across multiple institutions, achieving expert-level reproducibility. Our findings are consistent with these reports, as YOLOv8 achieved high accuracy and robustness despite being trained on a single-center dataset.

Previous work with YOLOv5-based networks has reported map values exceeding 80% for acute infarct detection [3,19], while Faster R-CNN has historically been considered the gold standard for accuracy in object detection tasks [4]. However, the slower inference speed of Faster R-CNN has limited its clinical applicability in time-sensitive scenarios. Al sufyani [4] showed that YOLOv8 surpassed Faster R-CNN in brain tumor detection, achieving higher recall and faster inference. Our study extends this evidence to stroke imaging, demonstrating that YOLOv8 not only maintains accuracy but also achieves inference times of under 15 MS per slice, three times faster than Faster R-CNN. In acute stroke workflows, where treatment decisions must be made within narrow therapeutic windows, such improvements in computational efficiency are of direct clinical relevance [8,9].

Benchmarking against radiologists remains essential for clinical translation. Prior meta-analyses have confirmed that CNNs can match radiologist-level performance for intracranial hemorrhage detection on CT [5], yet comparable evidence on MRI is sparse [6]. By directly comparing YOLOv8 and Faster R-CNN with radiologist consensus, our study provides novel evidence that AI systems can achieve sensitivity comparable to experts on DWI. Although specificity lagged behind radiologists, the narrowing gap highlights the growing potential of AI to complement human interpretation rather than replace it.

4.3 Clinical Implications

The clinical implications of these findings are substantial. First, YOLOv8 demonstrated performance that is clinically acceptable for use as a decision-support tool, particularly in settings where neuroradiology expertise is not readily available. In resource-limited environments, such systems could provide rapid screening and triage, reducing the risk of delayed diagnosis. Second, inference speed is a critical factor in acute stroke care. The ability of YOLOv8 to process an image in under 15 MS suggests that it could be integrated into near real-time workflows, such as PACS-based alerts or automated triage dashboards. Faster R-CNN, while accurate, is unlikely to meet the practical demands of stroke care due to its slower processing.

Furthermore, the integration of AI into clinical workflows has the potential to improve efficiency for radiologists. Automated lesion detection can serve as a “second reader,” highlighting suspicious areas and reducing fatigue-related errors, especially in high-volume settings. In stroke centers where time-to-treatment directly affects outcomes, such decision-support systems could contribute to improved patient survival and functional recovery [1,8,9].

4.4 Error Analysis

Error analysis revealed important patterns that warrant discussion. YOLOv8 occasionally misclassified susceptibility artifacts and cortical laminar necrosis as hemorrhage, while Faster R-CNN frequently failed to detect small lacunar infarcts. These misclassifications are not surprising, as they reflect challenges inherent in DWI interpretation. Similar issues have been reported in earlier studies, where small vessel disease and chronic ischemic changes often confounded AI models [12–14]. Addressing these errors will require multimodal integration, particularly with GRE and SWI sequences, which provide higher sensitivity for hemorrhage detection [6,10].

4.5 Limitations

This study has several limitations. First, only single representative slices were used for each case, rather than full volumetric datasets. This approach reduced computational complexity but may underestimate lesion burden and fail to capture the full extent of pathology. Second, the retrospective and single-center design limits generalizability, as the dataset may not reflect scanner variability or population heterogeneity seen in multicenter cohorts. Third, only DWI was analyzed; although DWI is highly sensitive for infarction, it is suboptimal for hemorrhage detection compared with GRE or SWI. Fourth, inference times were measured in a controlled research environment with high-performance GPUs, which may not reflect integration into standard hospital PACS systems. Finally, while radiologist benchmarks provided a strong reference, interobserver variability remains a factor despite consensus review.

4.6 Future Directions

Future work should address these limitations through multicenter, multi-sequence validation studies. Incorporating ADC, GRE, and SWI sequences alongside DWI could improve the detection of hemorrhagic lesions and reduce artifact misclassification. The use of explainability tools such as Grad-CAM would provide visual explanations of AI predictions, increasing clinician trust and aiding in error identification. Prospective trials are also necessary to evaluate the clinical impact of AI-assisted decision-making, including effects on time-to-treatment, diagnostic confidence, and patient outcomes.

In addition, federated learning approaches may enable collaborative model training across institutions without requiring data sharing, thereby overcoming barriers related to patient privacy and data governance [15–17]. Such strategies could substantially improve generalizability while addressing ethical concerns about centralized data storage.

5. CONCLUSION

This study demonstrated that automated detection of intracranial hemorrhage and acute infarction on diffusion-weighted MRI can be achieved with clinically meaningful accuracy using modern deep learning methods. YOLOv8, a one-stage object detector, consistently outperformed Faster R-CNN in lesion-level recall, image-level classification accuracy, and inference speed. While radiologists remained the gold standard, achieving near-perfect performance with sensitivity, specificity, and accuracy of 99%, YOLOv8 achieved sensitivity that approached expert level, with no significant statistical difference detected. Faster R-CNN, by comparison, showed lower sensitivity and slower inference, limiting its suitability for acute stroke workflows.

The clinical implications of these findings are important. In acute stroke, where every minute of delay reduces the likelihood of favorable recovery, models such as YOLOv8 could provide rapid triage support and assist non-specialist clinicians in differentiating infarction from hemorrhage. Real-time lesion detection has the potential to serve as a “second reader,” alerting radiologists to subtle abnormalities, reducing oversight, and ensuring consistency in high-volume or resource-limited settings. By complementing rather than replacing expert interpretation, AI systems may bridge the gap between limited specialist availability and the growing burden of stroke worldwide.

Looking ahead, further work must address the limitations of this study. Future efforts should incorporate multi-sequence and volumetric MRI data, validate performance across multicenter datasets, and integrate explainability features to enhance clinician trust. Prospective trials will be required to determine whether AI integration leads to improved treatment times, better patient outcomes, and broader healthcare equity. With careful validation and responsible deployment, YOLOv8 and similar one-stage detectors have the potential to transform acute stroke imaging, complementing radiologists and supporting faster, more reliable decision-making in one of medicine’s most time-critical emergencies.

Reference

- 1) Feigin VL, Brainin M, Norrving B, Martins S, Sacco RL, Hacke W, et al. World Stroke Organization (WSO): Global burden of stroke and risk factors in 2020 and progress since 1990. *Lancet Neurol*. 2021;20(10):795–820.
- 2) Bojsen JA, Elhakim MT, Graumann O, Gaist D, Nielsen M, Harbo FSG, et al. Artificial intelligence for MRI stroke detection: a systematic review and meta-analysis. *Insights Imaging*. 2024; 15:24.
- 3) Chen S, Duan J, Wang H, Wang R, Li J, Qi M, et al. Automatic detection of stroke lesion from diffusion-weighted imaging via the improved YOLOv5. *Comput Biol Med*. 2022; 150:106120.
- 4) Alsufyani A. Performance comparison of deep learning models for MRI-based brain tumor detection. *AIMS Bioeng*. 2025; 12:199–212.
- 5) Daugaard Jørgensen M, Antulov R, Hess S, Lysdahlgaard S. Convolutional neural network performance compared to radiologists in detecting intracranial hemorrhage from brain CT: a systematic review and meta-analysis. *Eur J Radiol*. 2022; 146:110073.

- 6) Luijten SPR, van der Ende NAM, Cornelissen SAP, Kluijtmans L, van Hattem A, Lycklama à Nijeholt G, et al. Comparison of DWI b0 with T2*-GRE or SWI for intracranial hemorrhage detection after reperfusion therapy for ischemic stroke. *Neuroradiology*. 2023;65(10):1649–55.
- 7) Khedr SA, Kassem HM, Hazzou AM, Awad E, Fouad MM. MRI diffusion-weighted imaging in intracranial hemorrhage. *Egypt J Radiol Nucl Med*. 2013;44(3):625–34.
- 8) Campbell BCV, Mitchell PJ, Kleinig TJ, Dewey HM, Churilov L, Yassi N, et al. Endovascular therapy for ischemic stroke with perfusion-imaging selection. *N Engl J Med*. 2015; 372:1009–18.
- 9) Goyal M, Menon BK, van Zwam WH, Dippel DWJ, Mitchell PJ, Demchuk AM, et al. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *Lancet*. 2016; 387:1723–31.
- 10) Ryu W-S, Schellingerhout D, Park J, Chung J, Jeong S-W, Gwak D-S, et al. Deep learning-based automatic segmentation of cerebral infarcts on diffusion MRI. *Sci Rep*. 2025; 15:13456.
- 11) Albers GW, Thijs VN, Wechsler L, Kemp S, Schlaug G, Skalabrin E, et al. Magnetic resonance imaging profiles predict clinical response to early reperfusion. *Ann Neurol*. 2006;60(5):508–17.
- 12) Lansberg MG, Thijs VN, Bammer R, Kemp S, Wijman CAC, Marks MP, et al. The MRA-DWI mismatch identifies patients with stroke who are likely to benefit from reperfusion. *Stroke*. 2007;38(2):470–76.
- 13) Bivard A, Levi C, Spratt N, Parsons M. Perfusion computed tomography: Imaging and clinical validation in acute ischaemic stroke. *Cerebrovasc Dis*. 2011;31(6):527–35.
- 14) Warach S, Chien D, Li W, Ronthal M, Edelman RR. Fast magnetic resonance diffusion-weighted imaging of acute human stroke. *Neurology*. 1992;42(9):1717–23.
- 15) Le Bihan D. Molecular diffusion, tissue microdynamics and microstructure. *Radiology*. 1990;177(2):328–29.
- 16) Beaulieu C. The basis of anisotropic water diffusion in the nervous system – a technical review. *NMR Biomed*. 2002;15(7–8):435–55.
- 17) Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *Br J Radiol*. 2019; 92:20180427.
- 18) Park J, Lee JS, Kang B, Lee H, Jeon JY, Kim HJ, et al. Artificial intelligence for stroke MRI segmentation: a multicenter validation study. *Stroke*. 2025;56(4):1457–66.
- 19) Cao H, Song Y, Li Y, Lin H, Qin J. YOLO-based automatic detection of cerebral infarction and hemorrhage using diffusion-weighted MRI. *Med Phys*. 2022;49(10):6521–33.