**Detection of Vascular Pathologies in Non-Contrast Magnetic Resonance Images for Early Screening**

**1. Abstract**

Accurate segmentation of ischemic stroke lesions in magnetic resonance imaging (MRI) is essential for early diagnosis and treatment planning. This study evaluates and compares the performance of three deep learning architectures—U-Net, nnU-Net, and YOLOv8—for the automatic segmentation of ischemic stroke lesions using FLAIR MRI sequences without contrast agents. The models were trained and tested on the ISLES 2022 public dataset, containing 250 patient cases. U-Net was manually configured as a 3D baseline model, nnU-Net was applied as a self-configuring 3D full-resolution mode, and YOLOv8 was adapted from object detection to 2D medical image segmentation. Performance was assessed using the Dice coefficient and Intersection over Union (IoU) metrics. Results show that nnU-Net achieved the best overall accuracy, with a mean Dice score of 0.483 ± 0.284, outperforming both U-Net and YOLOv8. YOLOv8 demonstrated the fastest inference and competitive segmentation accuracy, indicating potential for real-time clinical use. U-Net exhibited limited generalization across heterogeneous lesions. In conclusion, nnU-Net remains the most reliable and adaptive solution for MRI stroke lesion segmentation, whereas YOLOv8 offers a promising lightweight alternative for time-critical applications. These findings highlight that automated segmentation of ischemic lesions from non-contrast FLAIR MRI is feasible and may support faster, more objective, and clinically scalable stroke assessment.

**2. Introduction**

Ischemic stroke is one of the leading causes of disability and mortality worldwide, representing a major challenge for healthcare systems, with significant clinical and economic implications. A rapid and accurate diagnosis is crucial to initiate the appropriate treatment and to reduce long-term neurological deficits in patients [1]. In this context, medical imaging plays an essential role in the detection and delineation of brain regions affected by stroke. This process of identifying and outlining the affected regions is commonly referred to as image segmentation, a central task in medical image analysis. Among the different imaging modalities, magnetic resonance imaging (MRI) stands out as a key tool thanks to its ability to provide detailed anatomical and functional information without exposing patients to ionizing radiation.

The segmentation of ischemic lesions in MRI scans is a particularly demanding task. Traditionally, this process has relied on manual annotations by radiologists and other clinical experts. Although this approach is precise, it is also labor-intensive, time-consuming, and prone to inter-observer variability. These limitations hinder its applicability in large-scale clinical settings and delay the diagnosis of patients in acute or subacute stages. For this reason, the automation of lesion segmentation has been a long-standing goal in medical image analysis.

Classical image processing methods, including thresholding, region growing, clustering, and deformable models, have historically been applied in medical image segmentation. While these techniques can provide valuable insights, their performance is specially hindered by inherent challenges in medical imaging, such as anatomical variability among patients, low tissue contrast, imaging artifacts, and the high similarity between healthy and pathological regions. These limitations frequently result in inconsistent or suboptimal segmentation outcomes. Machine learning approaches, particularly support vector machines and random forests, represented a step forward by leveraging handcrafted features. However, these methods still rely heavily on feature engineering, which restricts their adaptability and generalization to new datasets.

In recent years, deep learning (DL) has revolutionized the field of medical image analysis by overcoming many of the limitations of traditional techniques [2]. Convolutional neural networks (CNNs) and their derivatives enable the automatic extraction of hierarchical features directly from raw image data, removing the need for manual feature design and significantly improving segmentation performance. Architectures such as U-Net and its variants have become the gold standard for biomedical image segmentation, demonstrating remarkable accuracy and robustness across multiple clinical applications. Moreover, DL-based methods provide consistency across many patients and can accelerate clinical workflows by reducing the burden on radiologists.

Despite these advances, the application of DL in ischemic stroke segmentation remains a complex challenge. Stroke lesions often exhibit heterogeneous appearances across patients, making them difficult to capture with a single segmentation strategy. Furthermore, the scarcity of annotated medical data adds an additional layer of difficulty. In this scenario, comparing different state-of-the-art architectures provides valuable insights into their strengths and weaknesses, as well as their potential for translation into real-world clinical environments.

The present research specifically focuses on the automatic segmentation of ischemic stroke lesions in FLAIR MRI sequences without the use of contrast agents. This choice is clinically relevant, as FLAIR images highlight hyperintense regions corresponding to edema and infarcted tissue, while avoiding the risks associated with contrast administration. The study explores and compares three distinct deep learning architectures:

* **U-Net**, a widely adopted baseline in medical image segmentation.
* **nnU-Net**, an advanced self-configuring framework that adapts its architecture and training strategy to the dataset at hand, consistently achieving state-of-the-art results in several medical image segmentation challenges.
* **YOLOv8**, a novel model originally designed for object detection, but recently extended to segmentation tasks, offering the opportunity to assess its applicability in the biomedical domain.

The specific objectives of this research are:

1. To implement and train U-Net, nnU-Net, and YOLOv8 architectures on a public dataset of ischemic stroke MRI scans.
2. To evaluate and compare their performance using standard segmentation metrics, including the Dice coefficient and Intersection over Union (IoU).
3. To analyze the strengths and limitations of each model in order to assess their potential clinical applicability.

The remainder of this article is structured as follows. Section 3 reviews the state of the art in medical image segmentation, focusing on both classical and deep learning approaches. Section 4 describes the dataset, preprocessing steps, and methodological details of the implemented models. Section 5 presents the experimental results, followed by a discussion in Section 6 that highlights the comparative analysis, limitations, and future directions. Finally, Section 7 summarizes the main conclusions.

**3. Related Work**

3.1 Scientific and Technical Context

While the clinical motivation for ischemic stroke lesion segmentation has been outlined in Section 1, it is equally important to situate this problem within the broader technical landscape of medical image analysis. Stroke lesion segmentation is among the most difficult challenges in brain imaging because of several specific characteristics: (i) lesions exhibit high inter-patient variability in size, shape, and location; (ii) tissue contrast in relevant MRI sequences such as FLAIR is often low, making lesions difficult to distinguish from surrounding white matter; (iii) infarcts evolve over time, so the appearance of acute, subacute, and chronic lesions differs substantially; and (iv) motion artifacts and scanner-dependent variations further complicate analysis. These aspects make the problem more complex than other established tasks such as brain tumor segmentation or multiple sclerosis lesion detection.

A major driver of technical progress has been the creation of benchmark challenges, in particular the ISLES series (Ischemic Stroke Lesion Segmentation) organized within the MICCAI community. These challenges have provided publicly available datasets, standardized evaluation metrics, and competitive leaderboards, enabling direct comparison across methods. Through ISLES, researchers have repeatedly confirmed that conventional intensity-based methods struggle, while data-driven and deep learning techniques consistently achieve superior performance.

3.2 Classification of Existing Approaches

Segmentation strategies for stroke imaging can be categorized into three major groups: traditional image processing methods, classical machine learning, and modern deep learning-based approaches.

Traditional image processing methods were the first to be applied to medical image segmentation. Thresholding techniques assign pixels to foreground or background based on intensity values, but they fail when tissue contrast is low. Region growing methods start from a seed point and iteratively aggregate neighboring pixels with similar intensity; however, they are sensitive to noise and initialization. Clustering approaches, such as k-means, group voxels according to statistical similarity, but lack spatial constraints and often oversegment complex structures. For instance, K-means has been applied in stroke lesion segmentation to identify hyperintense regions in brain images [3], demonstrating moderate success when combined with fuzzy clustering. Deformable models, including active contours (snakes) and level sets, introduce shape priors and regularization, allowing more flexible adaptation to object boundaries. Active contour models have been successfully applied in medical image segmentation, including brain scans, to delineate ischemic stroke lesions [4]. Similarly, statistical and deformable model approaches have been explored specifically for MRI-based ischemic stroke segmentation, highlighting their potential and limitations in handling heterogeneous lesion shapes [5]. Although these methods represented an advance over purely intensity-based techniques and provided valuable insights in early neuroimaging studies, they remain prone to convergence issues, require careful parameter tuning, and are rarely competitive against modern learning-based approaches.

Classical machine learning methods brought the use of data-driven classifiers trained on handcrafted features. Typical features included intensity histograms, texture descriptors, local gradients, or wavelet coefficients. Support vector machines (SVMs), random forests, and shallow neural networks were widely applied. For example, Geremia et al. [6]demonstrated the potential of random forests for lesion segmentation in multiple sclerosis, showing improved generalization compared to thresholding. In the context of ischemic stroke, random forests and SVMs have been applied to diffusion-weighted and FLAIR MRI to segment infarcted regions, achieving moderate accuracy [3]. These studies highlight that while classical machine learning approaches can improve segmentation over simple intensity-based methods, their performance is limited by the quality and representativeness of handcrafted features. Another limitation is that feature extraction pipelines are often dataset-specific, reducing transferability to new data. In addition, these models do not scale well to high-dimensional medical imaging datasets compared to modern deep learning approaches.

Deep learning methods have redefined the state of the art in medical image segmentation. Convolutional neural networks (CNNs) automatically learn hierarchical representations of the data, reducing the need for manual feature engineering. Fully convolutional networks (FCNs) introduced end-to-end learning for dense predictions. The U-Net architecture [7] became a landmark in biomedical image segmentation, with its encoder-decoder structure and skip connections enabling high-level contextual feature extraction and fine spatial localization, particularly relevant for small ischemic infarcts. Variants such as residual U-Nets, attention U-Nets [8], and multi-scale 3D U-Nets [9] have demonstrated improved performance by incorporating residual connections, attention mechanisms, and multi-modal MRI inputs (e.g., DWI and FLAIR) for stroke lesion segmentation.

Frameworks like nnU-Net [10] propose a fully automated strategy, self-configuring network depth, patch size, preprocessing, and training parameters to achieve robust performance across multiple medical imaging tasks, including ischemic stroke. Beyond U-Net derivatives, ensembles of fully convolutional networks have been applied to stroke and white matter hyperintensity segmentation, showing enhanced robustness and accuracy [11].

Other deep learning architectures originally developed for volumetric or general segmentation have also been adapted to medical imaging. Examples include DeepMedic [9]for multi-scale 3D CNNs with probabilistic inference, V-Net for volumetric segmentation, and 2D CNN-based frameworks for ischemic stroke lesion segmentation [12][13]. Additionally, general-purpose segmentation models such as SegNet and DeepLabv3+, initially designed for natural images, have been successfully applied to medical imaging tasks including stroke and brain lesion segmentation.

More recently, YOLO-based models have been explored for ischemic stroke lesion detection and segmentation, leveraging their speed and efficiency for real-time clinical applications. For example, MSA-YOLOv5 [14] incorporates multi-scale attention mechanisms to improve detection of small and low-contrast lesions in multimodal MRI images. Similarly, lightweight models such as TE-YOLOv5 [15]have demonstrated effective segmentation of diffusion-weighted imaging (DWI) stroke lesions while reducing computational requirements. These studies illustrate the potential of YOLO architectures to complement traditional 3D CNNs, offering rapid inference and promising accuracy, particularly in scenarios where real-time performance is critical.

These deep learning methods outperform traditional and classical machine learning approaches, especially in handling high-dimensional data and complex lesion variability. Nevertheless, they require large annotated datasets, substantial computational resources, and careful design choices or automated configuration strategies to generalize effectively across patients and imaging modalities.

3.4 Knowledge Gaps and Our Contribution

Several knowledge gaps remain in the current state of the art. First, there is a scarcity of large annotated ischemic stroke datasets, especially for FLAIR sequences, which limits the potential of deep learning models. Second, while U-Net and nnU-Net are widely studied, the applicability of emerging general-purpose segmentation models like YOLOv8 in the biomedical domain remains underexplored. Third, few systematic comparative studies have been conducted across models with fundamentally different design philosophies: U-Net as a domain-specific baseline, nnU-Net as a self-configuring medical framework, and YOLOv8 as a general-purpose real-time model.

The present study addresses these gaps by conducting a comparative analysis of U-Net, nnU-Net, and YOLOv8 for ischemic stroke lesion segmentation using FLAIR MRI. Although trained on a relatively small dataset, the results indicate that these models can achieve competitive performance. This suggests that even with limited annotated data, reliable automated segmentation is feasible. Furthermore, findings may be extrapolated to larger datasets in the future. By evaluating YOLOv8 alongside established medical architectures, this work contributes new insights into the potential of real-time segmentation methods in clinical imaging workflows, offering a perspective on how efficiency and accuracy can be balanced for practical applications.

**4. Methodology**

4.1 Dataset and Preprocessing

The experiments were conducted using the publicly available ISLES 2022 (Ischemic Stroke Lesion Segmentation) dataset, developed as part of the international MICCAI challenge. The dataset contains 250 multimodal MRI cases from patients diagnosed with ischemic stroke, each accompanied by manually annotated lesion masks created by expert radiologists. Every case includes several MRI sequences—Diffusion-Weighted Imaging (DWI), Apparent Diffusion Coefficient (ADC), and Fluid-Attenuated Inversion Recovery (FLAIR)—which capture complementary physiological information of the infarcted tissue.

For this study, only the FLAIR sequence was used as model input. FLAIR images suppress the cerebrospinal fluid (CSF) signal, enhancing the visibility of hyperintense regions that correspond to ischemic edema or infarction. This property makes FLAIR particularly suitable for the segmentation of small lesions, especially in subacute phases. Each image and its corresponding segmentation were provided in NIfTI (.nii) format, which preserves the spatial and volumetric information of 3D neuroimaging data. To ensure consistency between the MRI volumes and their corresponding segmentation masks, a general preprocessing step was applied prior to model-specific procedures. The segmentation masks provided in the ISLES dataset did not always match the spatial dimensions of their respective FLAIR images, which could hinder the application of automated segmentation models. Therefore, a spatial resampling process was carried out to align each mask with its corresponding MRI volume. This adjustment was performed using nearest-neighbor interpolation, which assigns to each new voxel the value of the closest voxel in the original mask, thus preserving the discrete nature of the labeled regions and preventing the generation of invalid intermediate values.

This preprocessing step was essential to guarantee the spatial coherence between input images and ground-truth segmentations, enabling consistent training and evaluation across all models. After this alignment, each model applied its own specific preprocessing pipeline, which is described in the corresponding subsections below.

4.2 Experimental Environment

All experiments were conducted using Python-based frameworks for deep learning and medical image processing. Training was performed on a GPU-equipped environment (NVIDIA Tesla T4) to ensure efficient execution of 3D convolutional networks. The implementations relied on TensorFlow/Keras for U-Net, nnU-Net for the self-configuring model, and the Ultralytics framework for YOLOv8. The setup ensured full reproducibility across experiments.

4.3 Model Architectures and Training Configuration

Three segmentation models were implemented and compared: U-Net, nnU-Net, and YOLOv8. These architectures were chosen to represent three complementary paradigms in medical image segmentation: a manually designed biomedical model, an automated self-configuring framework, and a general-purpose segmentation architecture adapted from object detection.

Each model was independently trained using the same dataset split, preprocessed FLAIR volumes, and evaluation metrics to ensure a fair comparison. Training configurations are summarized in Table 1.

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter** | **U-Net** | **nnU-Net** | **YOLOv8** |
| **Input Type** | 3D (192×192×64) | 3D full-resolution | 2D slices (reconstructed to 3D) |
| **Configuration Type** | Manual | Automatic (self-configuring) | Default (Ultralytics) |
| **Optimizer** | Adam (lr = 1e−4) | Default SGD (automatic) | Default (automatic) |
| **Loss Function** | BCE + Dice | Soft Dice + Cross Entropy (default) | Segmentation loss (default) |
| **Epochs** | 100 (stopped at 24, early stopping) | 141 (manual stop) | 100 (stopped at 50, manual stop) |
| **Batch Size** | 1 | Automatic (patch-based) | 8 |
| **Data Augmentation** | Flips, rotations (±15°) | Automatic (rotations, elastic deformations, brightness/contrast) | Flips, rotations, contrast |
| **Evaluation Metrics** | Dice | Dice | Precision, Recall, mAP; Dice (on test) |

Table 1. Training configuration for each model.

4.3.1 U-Net

The U-Net architecture follows a symmetric encoder–decoder design. The encoder progressively downsamples the input through convolutional and max-pooling layers to extract hierarchical features, while the decoder reconstructs the segmentation map through upsampling and skip connections that fuse spatial details from the encoder with the decoder’s learned representations. This structure enables precise localization, particularly for small and irregular ischemic lesions.

Before model training, all images underwent a standardized preprocessing pipeline to ensure consistency across the dataset. First, the volumes were spatially resampled to an isotropic resolution of 1×1×1 mm³ using linear interpolation for the MRI scans and nearest-neighbor interpolation for the segmentation masks to maintain their discrete labels. Then, each volume was intensity normalized using *z-score normalization* (subtracting the mean and dividing by the standard deviation) to standardize voxel intensities and facilitate model convergence. To ensure uniform input dimensions, all volumes were cropped or zero-padded to a fixed size of 192×192×64 voxels, centered around the brain region to preserve relevant anatomical information.

During training, data augmentation was applied to increase variability and mitigate overfitting. Random 3D flips along spatial axes and in-plane rotations of up to ±15° were performed with equal probability, always applied consistently to both images and masks. The dataset was divided into training (80%), validation (10%), and test (10%) subsets, ensuring that patient cases did not overlap across partitions.

The 3D U-Net was configured with convolutional blocks consisting of two 3×3×3 convolutions followed by batch normalization and ReLU activation. The encoder comprised four levels with 64, 128, 256, and 512 filters, and the bottleneck used 1024 filters. The decoder mirrored this configuration with deconvolutional layers followed by concatenation and convolutional refinement. The final layer was a 1×1×1 convolution with sigmoid activation, producing voxel-wise probabilities for binary segmentation.

Training was performed for up to 100 epochs with a batch size of 1 due to GPU memory constraints. The Adam optimizer was used with a learning rate of 1×10⁻⁴. The loss function combined Binary Cross-Entropy (BCE) and Dice Loss with equal weights (α = 0.5), balancing voxel-level accuracy and spatial overlap. The Dice coefficient was monitored as the main validation metric.

Regularization strategies included early stopping (patience = 20 epochs), model checkpointing (saving the best model based on validation Dice), and CSV logging for metric tracking. Data augmentation was applied on the fly during training to enhance robustness.

4.3.2 nnU-Net

The nnU-Net (no-new U-Net) framework represents an automated and self-configuring approach to biomedical image segmentation. Unlike manually tuned architectures, nnU-Net automatically determines preprocessing, patch size, normalization, and training parameters based on dataset characteristics.

For this study, the 3D full-resolution nnU-Net configuration was selected, which operates directly on volumetric data without downscaling. The preprocessing pipeline, model depth, and patch size were generated automatically by the framework. The network architecture retained the core principles of U-Net but adapted parameters such as kernel size, number of feature maps, and training schedule to the ISLES dataset.

Training was performed on a single fold of the default five-fold cross-validation scheme provided by nnU-Net, due to computational constraints. Although nnU-Net is designed to train and ensemble multiple folds for optimal generalization, this study focused on one representative fold to evaluate the feasibility and performance of the approach.

The training process followed the default nnU-Net configuration, which automatically applies a combination of Dice and Cross-Entropy loss and optimizes the network using stochastic gradient descent (SGD) with momentum.

During training, the framework also performed extensive on-the-fly data augmentation, including random spatial transformations (rotations, scaling, and elastic deformations) and intensity augmentations such as brightness and contrast adjustments. These operations were applied probabilistically to each patch, enhancing the model’s robustness to anatomical variability and scanner-related differences.

The model was trained for 141 epochs, at which point the training was manually stopped once convergence trends were observed. Despite the reduced training schedule, the results obtained were consistent with those reported in the literature, confirming the robustness of the nnU-Net framework even under limited computational resources.

4.3.3 YOLOv8

The YOLOv8-seg model extends the well-known object detection architecture into instance segmentation. It combines real-time efficiency with pixel-level prediction by integrating a convolutional backbone, a neck for multi-scale feature fusion, and segmentation heads that output binary masks for detected regions. Although originally designed for natural images, its adaptability and speed make it an interesting candidate for biomedical applications.

For adaptation to MRI data, the input images were converted into 2D slices extracted from the FLAIR volumes, maintaining spatial correspondence with their 3D masks. The model was initialized with random weights and trained specifically on the FLAIR images for binary ischemic lesion segmentation, without using any external pretrained parameters.

Training followed the default configuration provided by the Ultralytics YOLOv8-seg framework, which automatically defines optimization parameters and learning rate scheduling. The model was trained for 100 epochs with a batch size of 8, using the framework’s segmentation loss component as the primary objective.

During training, standard data augmentations were applied, including random flips, rotations, and contrast variations, to improve robustness and reduce overfitting.  
Model performance was monitored on the validation set using precision, recall, and mean Average Precision (mAP) metrics, as implemented by the YOLOv8 evaluation pipeline.  
After training, the Dice coefficient was computed on the independent test set to provide a direct comparison with the other segmentation models.

While YOLOv8 processes data in 2D, the resulting slice-wise masks were reassembled into volumetric segmentations during postprocessing, enabling fair comparison with the 3D U-Net and nnU-Net outputs.

4.4 Evaluation Metrics

Model performance was evaluated using the two most common overlap-based segmentation metrics: Dice Similarity Coefficient (DSC) and Intersection over Union (IoU).

The Dice coefficient quantifies the harmonic mean between precision and recall at the voxel level, providing a measure of overlap between predicted and reference masks. It is particularly suited for medical imaging tasks with imbalanced classes, such as small ischemic lesions. The IoU measures the ratio between the intersection and the union of predicted and ground-truth regions, offering a stricter assessment of segmentation accuracy.

Both metrics were computed on the test set for each model. Higher values indicate better segmentation performance, with 1 representing perfect overlap.

4.5 Ethical Considerations

The study was conducted entirely using publicly available and anonymized data, thus requiring no additional ethical approval. The ISLES 2022 dataset complies with the ethical and legal standards of medical data sharing and was collected following informed consent procedures by its original providers.

No personally identifiable information was processed, and all experiments were performed in compliance with the European General Data Protection Regulation (EGDPR) for research data handling. The study did not involve any clinical interventions or new patient data acquisition.

**5. Results**

This section presents the quantitative and qualitative results obtained from the three segmentation models: U-Net, nnU-Net, and YOLOv8. Performance metrics include Dice coefficient, Intersection over Union (IoU), and, in the case of YOLOv8, precision, recall, and mean Average Precision (mAP). Figures and tables illustrate representative outcomes and training behaviors.

5.1 U-Net

The 3D U-Net training was automatically stopped after 24 epochs due to the early stopping criterion, as shown in Figure 1, which depicts the training loss curve. The validation Dice coefficient increased steadily before reaching a stable value around epoch 20 (Figure 1). Quantitative results for the test set are summarized in Table 1.

Gráfico, Gráfico de líneas, Histograma

El contenido generado por IA puede ser incorrecto.

Figure 1. U-Net training metrics.

|  |  |
| --- | --- |
| **Metric** | **Mean ± SD** |
| **Dice** | 0.294 ± 0.282 |
| **IoU** | 0.207 ± 0.221 |

Table 2. Average performance of the U-Net on the test set.

Qualitative examples of the model outputs are presented in Figure 2, showing the ground truth in green and predicted segmentation masks in red, in both 2D slices and 3D reconstructions for representative test cases. The figure illustrates the best-performing case, with a Dice coefficient of 0.811, and an intermediate case, with a Dice of 0.342, highlighting the variability in segmentation accuracy across the dataset.

Un dibujo de una persona

El contenido generado por IA puede ser incorrecto.

Figure . U-Net representative examples.

5.2 nnU-Net

The nnU-Net model automatically configured its preprocessing, patch size, and training schedule. It reached optimal validation performance after 141 epochs, as indicated by the decreasing loss curve and the rise of the pseudo-Dice score, showed in Figure 3. The average results on the test set are displayed in Table 2.

Gráfico, Gráfico de líneas

El contenido generado por IA puede ser incorrecto.

Figure 3. nnU-Net training metrics.

|  |  |
| --- | --- |
| **Metric** | **Mean ± SD** |
| **Dice** | 0.483 ± 0.284 |
| **IoU** | 0.362 ± 0.244 |

Table 3. Average performance of the nnU-Net on the test set.

Visual results demonstrated that nnU-Net effectively captured lesion boundaries and spatial continuity, even in cases with irregular shapes or heterogeneous intensities. Figure 4 illustrates representative segmentations for two test cases (Dice 0.824 and 0.480), showing good agreement between the predicted and manual annotations.

Imagen que contiene foto, colorido, diferente, mostrar

El contenido generado por IA puede ser incorrecto.

Figure 4. nnU-Net representative examples.

5.3 YOLOv8

The YOLOv8 segmentation variant was trained on 2D FLAIR slices reconstructed into 3D volumes for evaluation. The loss curve showed rapid convergence after 50 epochs (Figure 5). Quantitative metrics for validation are summarized in Table 4, while the final test results are presented in Table 5.

Gráfico, Gráfico de líneas

El contenido generado por IA puede ser incorrecto.

Figure . YOLOv8 training metrics.

|  |  |
| --- | --- |
| **Metric** | **Mean ± DE** |
| **Precision** | 0.437 ± 0.057 |
| **Recall** | 0.213 ± 0.031 |
| **mAP@0.5** | 0.202 ± 0.034 |
| [**mAP@0.5:0.95**](mailto:mAP@0.5:0.95) | 0.073 ± 0.014 |

Table 4. Validation metrics for YOLOv8.

|  |  |
| --- | --- |
| **Metric** | **Mean ± DE** |
| **Dice** | 0.476 ± 0.287 |
| **IoU** | 0.358 ± 0.254 |

Table 5. Average performance of YOLOv8 on the test set.

Qualitatively, YOLOv8 achieved fast inference and detected most lesion regions, as shown in Figure 6, but its segmentations tended to be slightly less precise than those of the 3D networks, occasionally missing fine structures or small peripheral lesions.

Imagen que contiene foto, diferente, mostrar, colorido

El contenido generado por IA puede ser incorrecto.

Figure 6. YOLOv8 representative examples.

5.4 Summary of Comparative Performance

A comparative overview of the three architectures is provided in Table 6.

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Dice** | **IoU** | **Notes** |
| **U-Net** | 0.294 ± 0.282 | 0.207 ± 0.221 | Stable training; limited generalization to complex lesions |
| **nnU-Net** | 0.483 ± 0.284 | 0.362 ± 0.244 | Best overall performance and most consistent segmentation |
| **YOLOv8** | 0.476 ± 0,287 | 0.358 ± 0,254 | Fastest inference; lower boundary accuracy |

Table 6. Comparison of models on the ISLES 2022 test set.

As shown in Table 6, nnU-Net achieved the highest Dice and IoU values across all patient cases, followed by YOLOv8 and U-Net. These quantitative findings confirm the influence of model design and dimensionality on segmentation accuracy.

**6. Discussion**

The comparative analysis highlights distinct characteristics in the performance of U-Net, nnU-Net, and YOLOv8 for ischemic lesion segmentation in FLAIR MRI images.

The U-Net architecture achieved moderate performance across the validation and test sets. Despite its well-established effectiveness in biomedical image segmentation, the model exhibited considerable variability among individual cases. Quantitatively, the Dice and IoU metrics indicated that the network was capable of correctly identifying lesions in certain cases but failed to generalize consistently across the dataset. This uneven behavior suggests that the classical U-Net is highly sensitive to image quality and lesion contrast. In cases where boundaries were well defined, segmentation was accurate, while diffuse or low-contrast lesions often led to missed detections.

The relatively high standard deviation observed in the test results indicates limited robustness. This variability may stem from the restricted size of the dataset and from the intrinsic heterogeneity of ischemic lesions. Additionally, the classic U-Net requires extensive manual configuration, including input size definition, normalization strategy, and data augmentation parameters. These manual design choices can introduce performance fluctuations depending on how well they adapt to the dataset characteristics.

The nnU-Net outperformed the classical U-Net in both validation and test phases. Its self-configuring mechanism automatically selected preprocessing, normalization, and training hyperparameters, leading to higher Dice and IoU scores. The model demonstrated a strong ability to adapt to the dataset and segment complex or irregular lesions, confirming the effectiveness of automated optimization. However, some test cases still produced null segmentations, revealing that even advanced configurations can struggle with highly heterogeneous or low-quality images.

The improved consistency observed in nnU-Net reflects its capacity to generalize better from limited data. Nevertheless, the standard deviation across test results indicates residual instability, possibly due to the reduced number of samples and to contrast variations between scans. In addition, nnU-Net is computationally demanding, requiring long training times and significant GPU resources, which may limit its practical deployment in resource-constrained environments.

Although originally designed for object detection rather than medical segmentation, YOLOv8 achieved competitive results in this study. Its inference speed and implementation simplicity stand out as major advantages, making it suitable for real-time applications. The model achieved respectable Dice and IoU values, occasionally matching those of nnU-Net in favorable cases. However, it also exhibited variability, with some test cases showing incomplete or missing lesion predictions.

This behavior can be explained by YOLOv8’s design, which prioritizes detection speed and bounding-box optimization over pixel-level precision. Consequently, the model performs well when lesions are clearly delimited but tends to underestimate diffuse or low-contrast regions. Despite these limitations, YOLOv8 demonstrates strong potential for fast and flexible segmentation pipelines, especially if adapted with domain-specific augmentations or loss functions.

When comparing all three architectures, nnU-Net achieved the highest accuracy and stability, confirming the benefits of automatic configuration in medical image segmentation. U-Net, while less resource-intensive, showed limited generalization, particularly in challenging cases. YOLOv8, though less precise, offered the best balance between speed and usability. These results emphasize that performance depends not only on network complexity but also on the suitability of the architecture to the dataset characteristics.

However, several limitations must be acknowledged. The study was conducted on a relatively small dataset of 250 cases, using only FLAIR sequences, which restricts the models’ ability to exploit complementary multimodal information (e.g., DWI or ADC). Computational constraints limited full cross-validation in nnU-Net and reduced hyperparameter optimization in the other models. Moreover, the evaluation focused exclusively on segmentation accuracy without assessing the real clinical impact.

Future research could focus on incorporating multimodal MRI sequences (e.g., DWI and ADC) to improve lesion characterization and model generalization. In addition to classical data augmentation techniques, the creation of larger and more diverse datasets through the inclusion of new clinical cases and the integration of complementary public databases would be essential to achieve greater robustness. From a methodological perspective, combining 3D segmentation architectures with lightweight 2D frameworks like YOLOv8 could lead to hybrid models that preserve spatial precision while maintaining high computational efficiency. Such developments would contribute to the design of reliable and time-efficient automated systems for early ischemic stroke screening and diagnosis.

**7. Conclusions**

This study presented a comparative analysis of three deep learning architectures—U-Net, nnU-Net, and YOLOv8—for the automatic segmentation of ischemic stroke lesions in FLAIR MRI images. The main objective was to evaluate their performance, analyze their respective strengths and weaknesses, and assess their feasibility for clinical use. The results demonstrate that deep learning techniques can achieve accurate and reproducible segmentation of ischemic lesions even without the use of contrast-enhanced imaging, supporting the potential of artificial intelligence for early stroke detection and screening.

The experimental findings show that nnU-Net achieved the highest overall accuracy, obtaining the best Dice and IoU scores and exhibiting strong generalization across cases. Its self-configuring design proved particularly effective in adapting the model to the dataset characteristics, minimizing the need for manual tuning. U-Net, despite its widespread use as a baseline in biomedical segmentation, achieved the lowest performance among the evaluated models. Its results were less accurate and more variable, confirming the limitations of manually configured architectures when applied to heterogeneous clinical data. YOLOv8, originally designed for object detection, offered competitive performance and the fastest inference time, highlighting its potential for real-time or resource-constrained applications.

These outcomes indicate that no single architecture dominates in all aspects; rather, each offers complementary advantages. nnU-Net is best suited for research and diagnostic contexts requiring high accuracy, U-Net serves as a methodological baseline that provides a reference for evaluating more advanced models, and YOLOv8 enables rapid deployment in clinical workflows where processing speed is crucial. Together, they demonstrate that automated segmentation of ischemic lesions is technically feasible and can be integrated into radiological pipelines to support decision-making and reduce manual workload.

The achievement of accurate segmentation without contrast administration has significant implications for non-invasive and cost-efficient stroke assessment. Automated detection can assist radiologists in identifying subtle ischemic regions that might otherwise be overlooked, standardize the diagnostic process, and accelerate treatment planning. These models demonstrate strong potential for real-world implementation as assistive tools within hospital imaging systems or screening platforms. With further validation on larger and more diverse datasets, they could evolve into clinically deployable AI solutions capable of supporting early diagnosis, personalized monitoring, and potentially near real-time MRI analysis in emergency settings.

In conclusion, this work confirms that deep learning-based segmentation provides a promising pathway toward automated, precise, and accessible tools for ischemic stroke detection. By combining accuracy, adaptability, and computational efficiency, the explored architectures contribute to advancing medical imaging toward faster and more reliable diagnostic support. The integration of such AI-driven solutions into clinical practice could significantly enhance patient care, enabling earlier intervention and better outcomes in one of the world’s leading causes of disability and mortality.

**8. References**

[1] MedlinePlus, “Stroke,” https://medlineplus.gov/spanish/ency/article/000726.htm.

[2] X. Liu, L. Song, S. Liu, and Y. Zhang, “A review of deep-learning-based medical image segmentation methods,” *Sustainability (Switzerland)*, vol. 13, no. 3, 2021, doi: 10.3390/su13031224.

[3] M. B. S, M. B. S Assistant Professor, and A. T. Professor, “Brain Stroke Detection using Integrated K means and FCM in Brain Images,” 2018. [Online]. Available: https://www.researchgate.net/publication/353807519

[4] X. Qian, J. Wang, S. Guo, and Q. Li, “An active contour model for medical image segmentation with application to brain CT image,” *Med Phys*, vol. 40, no. 2, 2013, doi: 10.1118/1.4774359.

[5] B. Stein, D. Lisin, J. Horowitz, E. Riseman, and G. Whitten, “Statistical and deformable model approaches to the segmentation of MR imagery and volume estimation of stroke lesions,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, 2001, pp. 829–836. doi: 10.1007/3-540-45468-3\_99.

[6] E. Geremia, O. Clatz, B. H. Menze, E. Konukoglu, A. Criminisi, and N. Ayache, “Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images,” *Neuroimage*, vol. 57, no. 2, 2011, doi: 10.1016/j.neuroimage.2011.03.080.

[7] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” May 2015, [Online]. Available: http://arxiv.org/abs/1505.04597

[8] O. Oktay *et al.*, “Attention U-Net: Learning Where to Look for the Pancreas,” May 2018, [Online]. Available: http://arxiv.org/abs/1804.03999

[9] K. Kamnitsas *et al.*, “Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation,” *Med Image Anal*, vol. 36, 2017, doi: 10.1016/j.media.2016.10.004.

[10] F. Isensee *et al.*, “nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation,” in *Informatik aktuell*, 2019. doi: 10.1007/978-3-658-25326-4\_7.

[11] H. Li *et al.*, “Fully convolutional network ensembles for white matter hyperintensities segmentation in MR images,” *Neuroimage*, vol. 183, 2018, doi: 10.1016/j.neuroimage.2018.07.005.

[12] P. M. Shah *et al.*, “2D-CNN Based Segmentation of Ischemic Stroke Lesions in MRI Scans,” in *Communications in Computer and Information Science*, Springer Science and Business Media Deutschland GmbH, 2020, pp. 276–286. doi: 10.1007/978-3-030-63119-2\_23.

[13] Y. Gheibi, K. Shirini, S. N. Razavi, M. Farhoudi, and T. Samad-Soltani, “CNN-Res: deep learning framework for segmentation of acute ischemic stroke lesions on multimodal MRI images,” *BMC Med Inform Decis Mak*, vol. 23, no. 1, Dec. 2023, doi: 10.1186/s12911-023-02289-y.

[14] S. Chen *et al.*, “MSA-YOLOv5: Multi-scale attention-based YOLOv5 for automatic detection of acute ischemic stroke from multi-modality MRI images,” *Comput Biol Med*, vol. 165, 2023, doi: 10.1016/j.compbiomed.2023.107471.

[15] S. Chen *et al.*, “Automatic detection of stroke lesion from diffusion-weighted imaging via the improved YOLOv5,” *Comput Biol Med*, vol. 150, 2022, doi: 10.1016/j.compbiomed.2022.106120.

**9. Figures and Tables**

* Figure 1: U-Net training metrics (Archivo: \*Figure 1\*)
* Figure 2: U-Net representative examples (Archivo: \*Figure 2\*)
* Figure 3: nnU-Net training metrics (Archivo: \*Figure 3\*)
* Figure 4: nnU-Net representative examples (Archivo: \*Figure 4\*)
* Figure 5: YOLOv8 training metrics (Archivo: \*Figure 5\*)
* Figure 6: YOLOv8 representative examples (Archivo: \*Figure 6\*)
* Table 1: Training configuration for each model (Archivo: \*Table 1\*)
* Table 2: Average performance of the U-Net on the test set (Archivo: \*Table 2\*)
* Table 3: Average performance of the nnU-Net on the test set (Archivo: \*Table 3\*)
* Table 4: Validation metrics for YOLOv8 (Archivo: \*Table 4\*)
* Table 5: Average performance of the YOLOv8 on the test set (Archivo: \*Table 5\*)
* Table 6: Comparison of models on the ISLES 2022 test set (Archivo: \*Table 6\*)