engineering proceedings

MDPI

Proceeding Paper

Deep Learning for Ischemic Stroke Lesion Segmentation: A Comparative Study of Model Architectures †

Lyailya Cherikbayeva [1,2], Zarina Melis [1], Arman Yeleussinov [1,2], Aisulu Ataniyazova [1,2,*] and Aktoty Cherikbayeva [1]

[1] Department of Computer Science, Farabi University, 050040 Almaty, Kazakhstan; cherikbayeva.lyailya@gmail.com (L.C.); melis.zarina98@gmail.com (Z.M.); armankaznu@mail.ru (A.Y.); aktoty.cherikbaeva@mail.ru (A.C.)

[2] Laboratory of Artificial Intelligence and Robotics, Institute of Information and Computational Technologies, 050010 Almaty, Kazakhstan

* Correspondence: aisulu.ataniyazova@gmail.com

† Presented at the International Conference on Electronics, Engineering Physics and Earth Science (EEPES2026), Bandirma, Turkey, 24–27 June 2026.

Abstract

Accurate and automated segmentation of ischemic stroke lesions from magnetic resonance imaging (MRI) data is crucial for rapid clinical diagnosis, treatment planning, and patient prognosis. This study presents a comprehensive evaluation of various deep learning architectures for segmenting acute and subacute ischemic lesions using the multicenter ISLES 2022 dataset. We compare traditional 3D U-Net models, attention mechanisms, and modern hybrid architectures, specifically Swin-UNETR (Swin Transformer-based UNet TRansformer). Our experiments utilize multimodal MRI data, including diffusion-weighted images (DWI) and apparent diffusion coefficient (ADC) maps. Results demonstrate that the Swin-UNETR architecture, combined with a specialized pad or crop preprocessing pipeline, advanced data augmentation using the TorchIO library, and a combined Dice–Cross Entropy (DiceCE) loss function, provides competitive performance. The best Swin-UNETR model achieved a Dice of 0.717, precision of 0.793, recall of 0.708, and HD95 of 12.24 mm on the validation set, with a loss of 0.26. The baseline 3D U-Net achieved a comparable Dice of 0.726, while Swin-UNETR showed higher precision, recall, and lower HD95. This study highlights the effectiveness of integrating hierarchical transformers with convolutional decoders for extracting both local and global contextual information in complex medical imaging tasks.

Keywords: ischemic stroke; MRI segmentation; deep learning; Swin-UNETR; ISLES 2022; medical image analysis

Check for updates

Academic Editors: Teodor Iliev, Ivaylo Stoyanov, Grigor Mihaylov and Selahattin Kosunalp

Published: 9 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Ischemic stroke remains a leading cause of long-term disability and mortality worldwide. Rapid identification and quantification of the infarct core are essential for determining patient suitability for reperfusion therapy, such as intravenous thrombolysis or mechanical thrombectomy. Magnetic resonance imaging (MRI), particularly diffusion-weighted imaging (DWI) and associated apparent diffusion coefficient (ADC) maps, is the gold standard for detecting acute ischemic changes due to its high sensitivity to water diffusion restriction in the brain.

Despite its clinical significance, manual segmentation of stroke lesions is a labor-intensive process subject to operator bias. Automated segmentation methods have the

potential to provide objective, reproducible, and rapid assessment. However, this task is challenging due to the high variability in lesion size, shape, and location, as well as the presence of imaging artifacts and a pronounced class imbalance between lesioned and healthy brain tissue. Recent advances in deep learning, particularly convolutional neural networks (CNNs) such as 3D U-Net, have set new standards in medical image segmentation. More recently, the advent of Vision Transformers (ViTs) has enabled the modeling of long-term spatial dependencies that are often missed by the local receptive fields of standard convolutions. In this paper, we examine the performance of several modern architectures, including 3D U-Net, Attention U-Net, ResNet-based models, and the hybrid Swin-UNETR, for ischemic stroke lesion segmentation using the ISLES 2022 dataset.

2. Literature Review

The field of medical image segmentation has evolved from atlas-based, thresholding- and rule-based methods to deep learning architecture based on the encoder–decoder paradigm. The key starting point was U-Net, which proposed a symmetrical structure with bandwidth connections that allowed for simultaneous consideration of local granularity and high-level context [1]. For three-dimensional medical data, this idea was developed in 3D U-Net, which became one of the basic volumetric segmentation models and set the standard for working with MRI volumes [2]. Later, the nnU-Net self-configuring framework showed that a significant part of the success in medical segmentation is determined not only by the architecture itself, but also by the automated selection of preprocessing, patch sampling, training schedule and post-processing [3]. Extensions of the U-Net family have increasingly incorporated attention mechanisms. For example, the study by [4] employed transformer-based architectures with attention mechanisms for 3D ischemic stroke segmentation in non-contrast CT, demonstrating improved feature extraction and segmentation performance in complex medical imaging tasks. In parallel, the residual connections introduced in ResNet were integrated into U-Net derivatives and SegResNet-like models to stabilize the learning of deeper networks and reduce gradient degradation [5]. However, purely CNN approaches still have a fundamental limitation: the local nature of the convolutions makes it difficult to model long-range spatial dependencies and global anatomical structure. That is why transformer architecture has appeared in medical segmentation, in particular UNETR and Swin-UNETR, where the encoder is based on self-attention and hierarchical feature extraction, and Swin Transformer additionally reduces the computational cost of global attention due to the shifted-window strategy [6].

In the tasks of segmentation of ischemic stroke, this evolution is particularly noticeable, since foci exhibit high variability in size, shape, number, and contrast at different MRI modalities. The ISLES series competitions played the role of a key comparative testing ground in this area: ISLES 2015 formed an early benchmark for multispectral stroke MRI, and ISLES 2022 consolidated a multicenter setting with DWI/ADC/FLAIR and hidden-test evaluation focused on real generalizability of models [7,8]. Against this background, challenge-derived systems have become not just competitive solutions, but an empirical base for clinical validation. In De la Rosa et al., DeepISLES has shown that the best ISLES 2022 models retain clinical usefulness outside the leaderboard scenario and are able to achieve stable results in a more realistic setting [9]. At the same time, related areas were developing: ensembles and transfer learning for inter-center sustainability [10], hybrid U-Net-Transformer architectures with self-supervised components [11], transformer-augmented MRI analysis to improve the quality of segmentation and take into account the global context [12], as well as registration-aware frameworks in which segmentation integration and intermodal alignment is considered as a single task [13].

https://doi.org/10.3390/engproc2026154067

The integration of multimodal data, primarily DWI, ADC, and FLAIR, remains critically important for accurate identification of the ischemic focus, since different sequences carry complementary information about diffusion restriction, tissue structure, and lesion stage. [8,13]. At the same time, multimodality dramatically complicates the feature space and makes the model more sensitive to misregistration, differences in scanning protocols, and cross-site domain shift. For this reason, either attention- and loss-aware CNN modifications are becoming increasingly important, for example, Attention U-Net with generalized Dice focal loss, focused on more reliable allocation of small foci [14], or semi-supervised and cross-modal strategies, which make it possible to use a limited amount of complete markup more efficiently [15]. Modern systematic reviews show that the field is rapidly shifting towards clinically validated, robust, and multimodal deep learning systems, but they do not confirm the existence of a universally best architecture for all stroke MRI segmentation scenarios. The most reasonable current conclusion is that transformer and Swin-oriented models are promising primarily as components of hybrid systems, whereas strong 3D U-Net, nnU-Net, and residual/SegResNet approaches continue to set a high and often more reproducible quality standard, especially in limited samples and in conditions of rigorous clinical validation [3,9].

3. Materials and Methods

3.1. Dataset: ISLES 2022

The study uses the ISLES 2022 dataset [9], a multicenter collection of MRI cases of patients with acute and subacute ischemic stroke. The dataset includes a total of 250 MRI cases from various equipment manufacturers. All images are provided in the NIfTI format (.nii.gz), which is a standard for 3D medical imaging, preserving spatial metadata and voxel sizes. For each patient, we use two main modalities: DWI (diffusion-weighted imaging): highly sensitive to acute ischemia; and ADC (adjusted diffusion coefficient): provides a quantitative measure of diffusion. Ground truth masks were created through a rigorous manual segmentation process. Experienced neuroradiologists manually delineated infarct boundaries on the DWI and ADC maps. To ensure high-quality annotations and minimize inter-operator variability, a multi-stage validation process was used, during which initial segmentations were reviewed and refined by senior experts. This manual process is considered the "gold standard", but highlights the complexity of the task due to the unclear boundaries and variable intensities of stroke lesions (Figure 1).

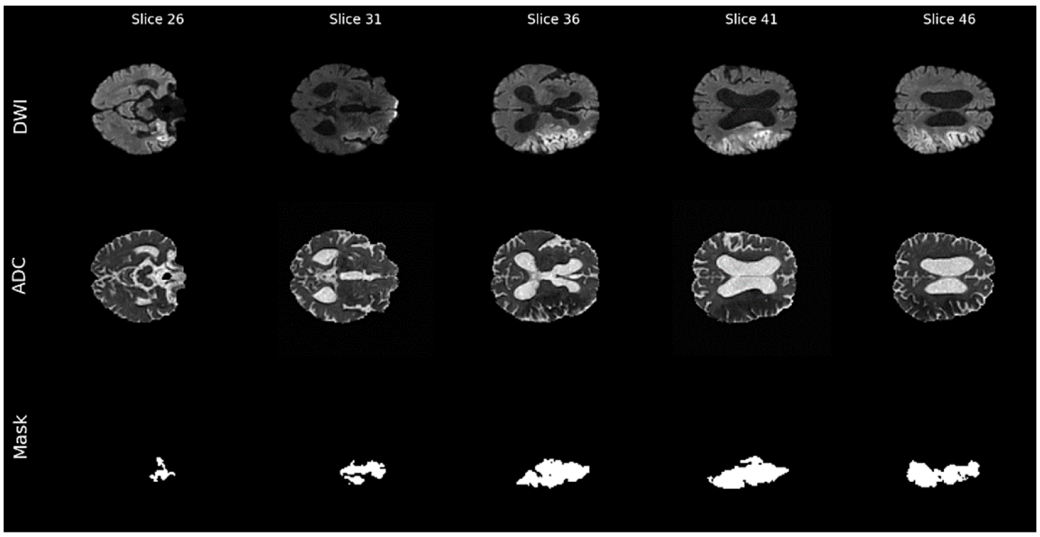


Figure 1. Multimodal axial MRI slices of the brain (DWI and ADC) with corresponding binary masks of ischemic lesions.

Figure 1 shows examples of axial brain MRI slices (DWI and ADC), along with the corresponding ischemic lesion masks. The hyperintensities in DWI and hypointensities in ADC are consistent with the labeled stroke zones, confirming the correctness of the annotations and the suitability of the data for training segmentation models.

3.2. Volumetric Preprocessing and Standardization

3.2.1. Spatial Standardization (Padding and Cropping)

Given the variability of input dimensions across MRI scanners, all 3D volumes were standardized to a fixed spatial resolution of 128 × 128 × 96 voxels. This standardization was achieved using a two-step spatial transformation:

1. Zero-Padding: If the input volume dimensions D_{in}, H_{in}, and W_{in} are smaller than the target dimensions (D_{target}, H_{target}, and W_{target}), the volume is symmetrically padded with zeros until the target size is reached.
2. Center Cropping: If the input dimensions exceed the target, center cropping is applied to extract the central region of interest. This ensures that the anatomical core of the brain is preserved while maintaining a consistent input size for the hierarchical transformer blocks.

3.2.2. Intensity Normalization

To minimize the impact of different signal-to-noise ratios and intensity scales across manufacturers, voxel intensities for DWI and ADC modalities were normalized using the global Z_{score} transform. For each volume, the normalized intensity, I_{norm}, is calculated as:

$$I_{norm} = \frac{I - \mu}{\sigma - \epsilon} \quad (1)$$

where I is the original voxel intensity, μ is the average volume intensity, σ is the standard deviation, and $\epsilon = 10^{-6}$ is a small constant added for numerical stability.

This ensures that the input features follow a standard normal distribution, which promotes faster convergence during the training phase.

3.2.3. Mask Binarization

The expert masks from the reference segmentation were binarized to create a clear target variable. Binarization is performed according to the following rule:

$$M(x,y,z) = \begin{cases} 1, \ if \ \ M(x,y,z) \ > \ 0 \\ 0 \end{cases} \quad (2)$$

This transformation otherwise maps any non-zero annotation to a single "focal" class, effectively formulating the problem as a binary volumetric segmentation problem.

3.3. Data Augmentation

3.3.1. Geometric Transformations

Geometric variability is provided by a combination of spatial transformations: Mirroring.

Random horizontal reflection along the sagittal plane is applied with a probability of $p = 0.5$, accounting for the bilateral symmetry of the brain. Affine transformations: Random affine transformations are used to model variations in patient position and the scale of anatomical structures. They include isotropic scaling in the range [0.9, 1.1], rotations of up to ±5° in all three planes, and translations of up to ±3 voxels, with a probability of $p = 0.5$.

https://doi.org/10.3390/engproc2026154067

3.3.2. Simulation of Intensity and Artifacts

To improve robustness to signal variations and MRI artifacts, the following intensity transformations are used: Additive Gaussian Noise—Gaussian noise with zero mean and $\sigma = 0.05$ ($p = 0.5$) is added to the voxel intensities, simulating signal acquisition noise. Bias Field—A spatially variable bias field is modeled to simulate intensity distortions associated with magnetic field inhomogeneity and sensitivity ($p = 0.3$), which is especially relevant for multicenter data.

3.4. Hybrid Hierarchical Architecture: Swin-UNETR

The primary architecture used in this study is Swin-UNETR (Figure 2), a hybrid model that combines the global modeling capabilities of hierarchical transformers with the localized feature extraction of convolutional decoders.

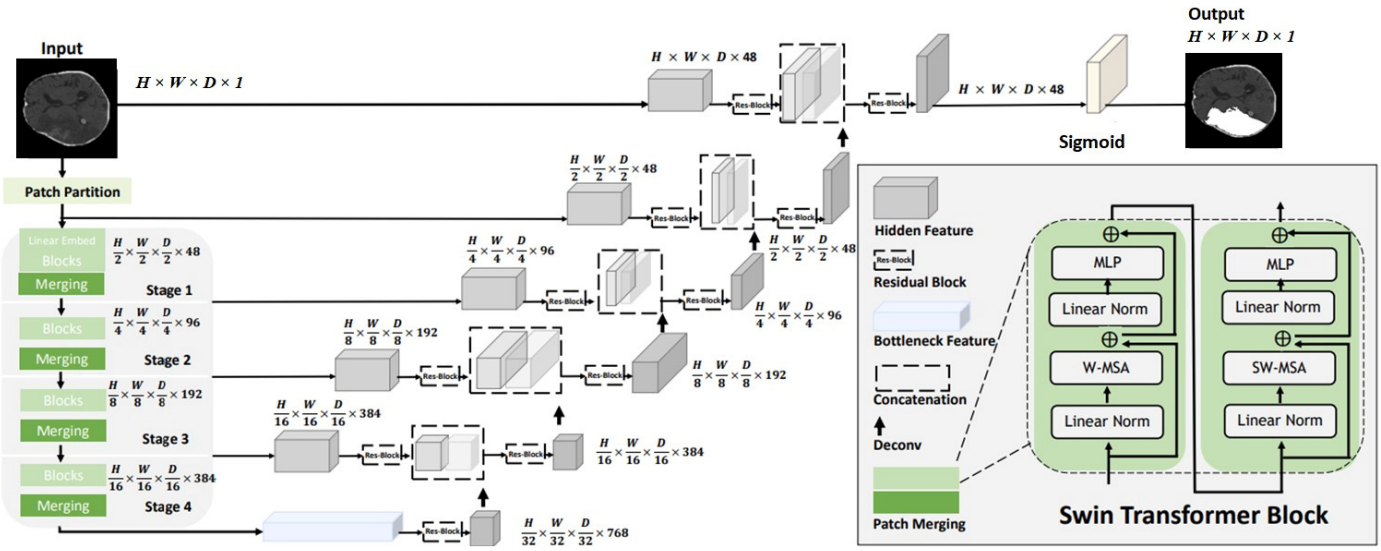


Figure 2. Swin UNETR model architecture.

The encoder is based on a hierarchical Swin Transformer, which processes the input volume at multiple scales. Unlike standard transformers, which compute global self-attention, the Swin encoder uses a Shifted Window-based Self-Attention (W-MSA) mechanism. This approach restricts attention computation to local windows while allowing connections between windows via cyclic shifts, significantly reducing the computational complexity from $O(N^2)$ to $O(N)$. The encoder configuration includes:

Feature Embedding: initial embedding layer with feature size C = 24.

Hierarchical stages: four stages with depths of (2, 2, 2, 2) and multi-head attention configurations of (3, 6, 12, 24) heads, respectively.

Normalization: group or instance normalization is applied at each stage to ensure a stable feature distribution across different MRI modalities.

The decoder follows a traditional U-shaped structure, using 3D convolutional blocks to upsample feature maps and restore spatial resolution. Hierarchical features from the transform encoder are combined with decoder features via skip connections, ensuring the preservation of fine-grained spatial information. The final output layer uses a $1 \times 1 \times 1$ convolution followed by a sigmoid activation function to produce a volumetric probability map.

https://doi.org/10.3390/engproc2026154067

3.5. Optimization and Loss Function

The model is optimized using the AdamW algorithm with a learning rate of $\eta = 10^{-4}$ and a weight decay of $\lambda = 10^{-5}$. A dynamic learning rate scheduler is used, which reduces the learning rate by a factor of 0.5 if the validation Dice coefficient does not improve for seven consecutive epochs. The objective function is a combined Dice–Cross Entropy (DiceCE) loss, which balances the overlap-based Dice loss and the pixel-wise cross entropy. The overall loss, L_{total}, is defined as:

$$L_{total} = L_{Dice} + L_{CE} \quad (3)$$

where L_{Dice} ensures reliable segmentation of focal boundaries, and L_{CE} ensures stable gradients for the overall intensity distribution.

4. Experiments and Results

The dataset was split into training (80%) and validation (20%) sets using a fixed random seed (42). Due to the limited dataset size, no separate test set was held out; the validation set was used for the final model evaluation. Training was conducted over 200 epochs with a batch size of two.

The quantitative results of the experimental series are presented in Table 1. Thirteen different configurations were analyzed, including variations in input modalities, model architectures, and data preprocessing strategies.

Table 1. Performance comparison of different models and configurations.

No.	Modality	Model	Loss Function	Train Dice	Train Loss	Val Dice	Val Loss
1	DWI	3D U-Net	DiceBCELoss	0.86	0.14	0.72	0.29
2	DWI + ADC	3D U-Net	DiceBCELoss	0.83	0.27	0.70	0.45
3	DWI + ADC	3D U-Net + Augment	DiceBCELoss	0.71	0.31	0.66	0.35
4	DWI + ADC	3D U-Net + Augment_v2	DiceBCELoss	0.74	0.28	0.69	0.34
5	DWI + ADC	3D U-Net + Patch	DiceBCELoss	0.80	0.28	0.67	0.55
6	DWI + ADC	ResNet + Augment_v2	DiceBCELoss	0.69	0.58	0.66	0.67
7	DWI + ADC	Attention U-Net + Augment_v2	DiceBCELoss	0.77	0.25	0.69	0.32
8	DWI + ADC	Attention U-Net + Augment_v2 + Patch	DiceBCELoss	0.71	0.38	0.70	0.63
9	DWI + ADC	Swin-UNETR + Augment_v2	DiceBCELoss	0.75	0.27	0.68	0.35
10	DWI + ADC	Swin-UNETR + Augment_v2	DiceFocalLoss	0.73	0.30	0.65	0.37
11	DWI + ADC	Swin-UNETR + Augment_v2 + Pad/Crop	DiceCELoss	0.80	0.18	0.71	0.26
12	DWI + ADC	Swin-UNETR + Augment_v2 + Patch	DiceBCEFocalLos	0.75	0.28	0.68	0.36
13	DWI + ADC	Swin-UNETR + Augment_v2 + Patch	DiceCELoss	0.77	0.21	0.69	0.29

As shown in Table 1, the Swin-UNETR model (Configuration 11) demonstrated the most stable and robust performance on the validation set. Although Configuration 1 (3D U-Net trained exclusively on DWI) demonstrated high metric values on the training set, it showed signs of overfitting compared to the hybrid transformer-based architecture. To better understand the differences, we computed additional metrics (precision, recall, and Hausdorff distance (HD95)) on the same validation set. Table 2 presents a side-by-side comparison of Configurations 1 and 11. The baseline 3D U-Net edges out Swin-UNETR slightly on Dice (0.72 vs. 0.717), but this difference is marginal. More importantly, Swin-UNETR surpasses the baseline in precision (+5.2%), recall (+4.1%), and HD95 (−12.8%), meaning it delineates lesion boundaries more accurately and produces fewer false positives. The validation loss of Swin-UNETR (0.26) is also lower than that of 3D U-Net (0.29), indicating better-calibrated probability estimates.

Table 2. Comparison of 3D U-Net and Swin-UNETR on the validation set.

No.	Model	Dice	Precision	Recall	HD95	Loss
1	3D U-Net	0.726	0.754	0.680	14.04	0.29
11	Swin-UNETR + Augment_v2 + Pad/Crop	0.717	0.793	0.708	12.24	0.26

The integration of the Pad/Crop spatial standardization strategy played a significant role in improving the model's robustness. This strategy ensured consistent input dimensions and helped stabilize the Swin Transformer self-attention mechanism. As a result, a minimal validation loss function was achieved (Val Loss = 0.26). The experimental results allow us to formulate the following key findings:

Superiority of Swin-UNETR with Pad/Crop: The Swin-UNETR model (Exp. 11) achieved the best balance between training and validation performance, achieving a Val Dice of 0.71 and a minimal Val Loss of 0.26.

Impact of Preprocessing: the custom pad-or-crop logic was crucial for standardizing the input data for the Swin Transformer blocks, which are sensitive to spatial dimensions.

Augmentation Performance: using TorchIO affine transformations and bias field augmentation helped the model better generalize to multicenter variations in the ISLES 2022 dataset.

To assess the contribution of each MRI modality, an ablation study was performed by training separate Swin-UNETR models on DWI only, ADC only, and their combination (DWI + ADC). The combined model corresponds to Configuration 11 from Table 1. Table 3 summarizes the results. The single-modality models underperform compared to the full combination. The DWI-only model reaches a Dice of 0.703. The ADC-only model does worse—0.318. This is expected, since ADC maps are derived from DWI and are rarely interpreted alone in clinical practice. Their role is to add complementary context, not to serve as a standalone sequence. When the two are combined, clear improvements are seen across all metrics. The full model (Config 11) achieves a Dice of 0.717, precision of 0.793, recall of 0.708, and HD95 of 12.24 mm. These findings confirm that DWI and ADC carry complementary information, and their combination is necessary for accurate stroke lesion segmentation.

Table 3. Ablation study of input modalities for Swin-UNETR on the validation set.

Modality	Model	Dice	Precision	Recall	HD95	Loss
DWI only	Swin-UNETR + Augment_v2 + Pad/Crop	0.703	0.781	0.683	12.65	0.28
ADC only		0.318	0.397	0.338	37.13	0.65
DWI + ADC		0.717	0.793	0.708	12.24	0.26

The qualitative performance of the best model (Swin-UNETR + Augment_v2 + Pad/Crop) is visualized in Figure 3. The figure shows three representative ischemic stroke cases, demonstrating the input modalities (DWI and ADC), expert annotation (True Mask), automatic model prediction, and the resulting overlay (Overlay) on the anatomical scan.

As shown in Figure 3, the model demonstrates high spatial accuracy in identifying infarct cores for lesions of various sizes and locations. Predictions closely match expert annotations, effectively capturing the irregular boundaries of ischemic lesions. Integration of DWI and ADC modalities allows the model to distinguish acute ischemic changes from other pathologies, while hierarchical transformer blocks ensure preservation of global anatomical context, minimizing false positives in healthy brain tissue.

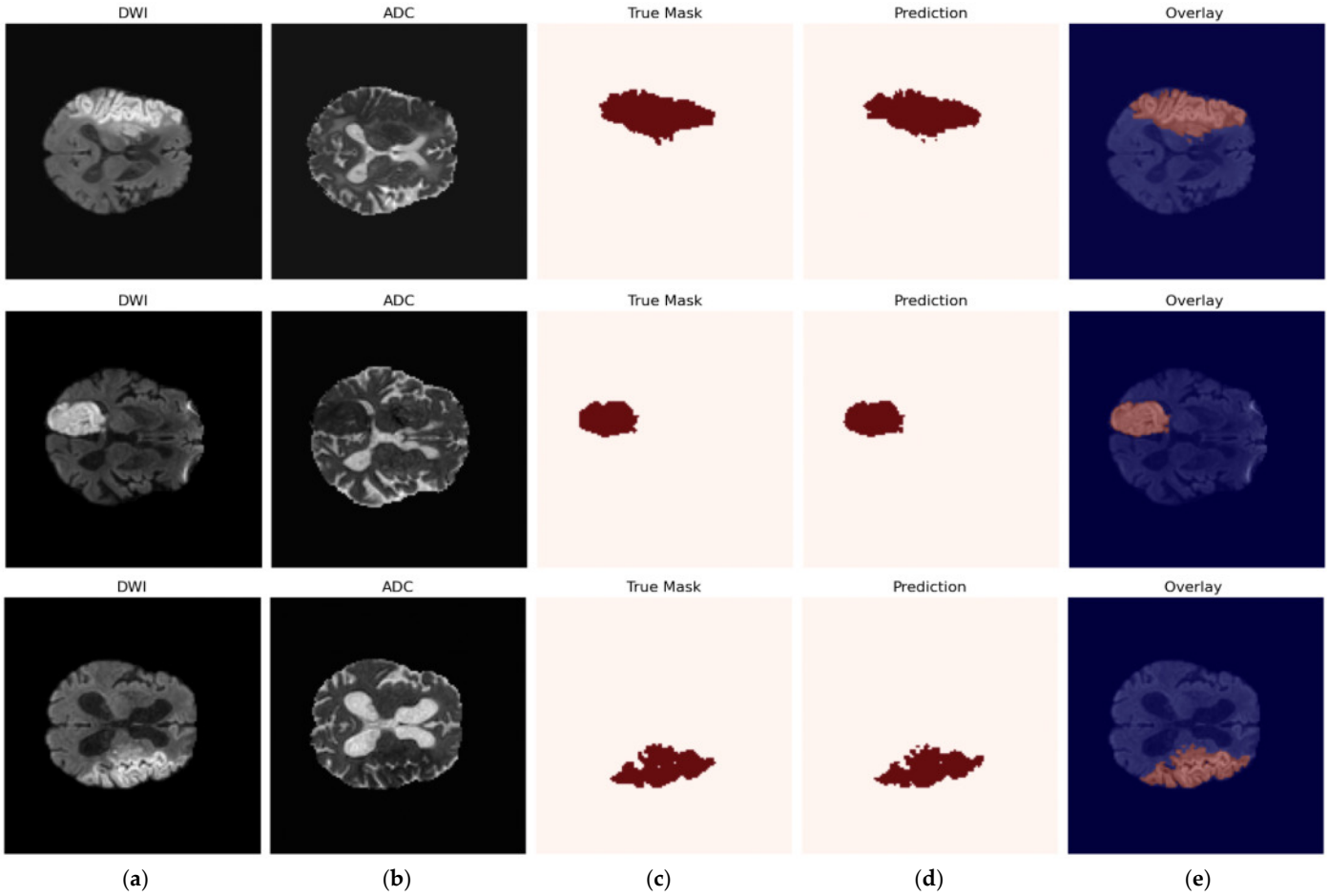


Figure 3. Visualization of Swin-UNETR model predictions: (**a**) DWI; (**b**) ADC; (**c**) True Mask; (**d**) prediction; (**e**) Overlay.

5. Discussion

The results presented above tell a more nuanced story than a simple comparison of Dice scores. How the two architectures compare: On the surface, Swin-UNETR and 3D U-Net deliver similar Dice values 0.717 versus 0.726. A closer look at other metrics reveals where the differences actually lie. Swin-UNETR achieves higher precision (0.793 vs. 0.754) and a lower Hausdorff distance (12.24 mm vs. 14.04 mm). What this means in practical terms is fewer false positives and boundaries that stick closer to what experts drew. For a clinician, these differences matter more than a tiny gap in Dice.

Why Swin-UNETR handles boundaries better: The self-attention mechanism in Swin-UNETR looks at the whole image at once, not just small local patches like convolutional layers do. This global view helps the model figure out where a lesion cannot possibly be, reducing false alarms outside the true stroke territory. Large or oddly shaped lesions benefit the most from this long-range awareness.

Why preprocessing made a difference: The pad-or-crop step turned out to be critical. Transformer-based models expect consistent input sizes; their attention windows and positional encodings depend on it. Simple resizing would have stretched or squeezed the voxels, distorting distances and intensities. Padding and cropping avoid this problem by keeping the original data intact. For a dataset like ISLES 2022, where scanners produce volumes of different sizes, this matters a great deal.

Generalization and overfitting: The training logs show that 3D U-Net drops from 0.86 on training data to 0.72 on validation, a gap of 0.14. Swin-UNETR drops from 0.80 to 0.71, a gap of only 0.09. The smaller gap suggests that Swin-UNETR learns features that

are less tied to the specific training examples and more transferable to unseen data. This is likely another benefit of the attention mechanism, which focuses on global patterns rather than local textures that might vary across different MRI scanners.

What each modality contributes: The ablation study (Table 3) trained three separate models—one on DWI alone, one on ADC alone, and one on both. DWI by itself reached a Dice of 0.703. ADC alone did worse at 0.318. That is not surprising: ADC maps come from DWI and are rarely interpreted on their own in real clinical work. They exist to add context, not to stand alone. Put the two together, and the model climbs to 0.717. The takeaway is clear, DWI and ADC complement each other, and you need both for the best results.

Limitations of the study: A few caveats deserve mention. First, the dataset has only 250 cases, which is modest by deep learning standards. No separate test set could be reserved, the validation set (20% of the data) had to serve as the final benchmark. This is common practice in small-sample medical imaging studies, but it is still a limitation. Second, Swin-UNETR demands more computation than 3D U-Net, which could be an issue in hospitals with limited GPU resources. Future work should test the model on external datasets and develop interpretability tools that help clinicians understand why the model makes a given prediction.

6. Conclusions

In this study, we conducted a rigorous comparative study of deep learning models for ischemic stroke lesion segmentation using the ISLES 2022 dataset. Our results demonstrate that the Swin-UNETR architecture, combined with data augmentation using TorchIO v.0.21.3, Pad/Crop preprocessing, and the DiceCE loss function, provides the most accurate and robust segmentation. These results provide a solid foundation for the development of automated clinical tools to assist radiologists in the rapid and accurate assessment of ischemic stroke.

Author Contributions: Conceptualization, L.C.; methodology, A.Y. and Z.M.; software, Z.M. and A.A.; validation, Z.M. and A.Y.; formal analysis, L.C. and A.A.; investigation, Z.M., A.Y. and A.C.; data curation, Z.M.; writing—original draft preparation, Z.M., A.Y. and A.A.; writing—review and editing, A.A. and Z.M.; visualization, A.C.; supervision, L.C.; project administration, L.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (AP26195405).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author.

Acknowledgments: During the preparation of this manuscript the authors used OpenAI ChatGPT, version GPT-5.3, for the purposes of language editing, structural suggestions and improvement of clarity. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015; pp. 234–241. [CrossRef]
2. Çiçek, Ö.; Abdulkadir, A.; Lienkamp, S.S.; Brox, T.; Ronneberger, O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016, Athens, Greece, 17–21 October 2016; pp. 424–432. [CrossRef]
3. Isensee, F.; Jaeger, P.F.; Kohl, S.A.A.; Petersen, J.; Maier-Hein, K.H. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **2021**, *18*, 203–211. [CrossRef] [PubMed]
4. Cherikbayeva, L.; Berikov, V.; Melis, Z.; Yeleussinov, A.; Baigozhanova, D.; Tasbolatuly, N.; Temirbekova, Z.; Mikhailapov, D. Ensembling Transformer-Based Models for 3D Ischemic Stroke Segmentation in Non-Contrast CT. *Appl. Sci.* **2025**, *15*, 9725. [CrossRef]
5. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
6. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 9992–10002. [CrossRef]
7. Maier, O.; Menze, B.H.; von der Gablentz, J.; Häni, L.; Heinrich, M.P.; Liebrand, M.; Winzeck, S.; Basit, A.; Bentley, P.; Chen, L.; et al. ISLES 2015—A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. *Med. Image Anal.* **2017**, *35*, 250–269. [CrossRef] [PubMed]
8. Hernandez Petzsche, M.R.; de la Rosa, E.; Hanning, U.; Wiest, R.; Valenzuela, W.; Reyes, M.; Meyer, M.; Liew, S.-L.; Kofler, F.; Ezhov, I.; et al. ISLES 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset. *Sci. Data* **2022**, *9*, 762. [CrossRef] [PubMed]
9. de la Rosa, E.; Reyes, M.; Liew, S.L.; Hutton, A.; Wiest, R.; Kaesmacher, J.; Hanning, U.; Hakim, A.; Zubal, R.; Valenzuela, W.; et al. DeepISLES: A clinically validated ischemic stroke segmentation model from the ISLES'22 challenge. *Nat. Commun.* **2025**, *16*, 7357. [CrossRef] [PubMed]
10. Jeong, H.; Lim, H.; Yoon, C.; Won, J.; Lee, G.Y.; de la Rosa, E.; Kirschke, J.S.; Kim, B.; Kim, N.; Kim, C. Robust Ensemble of Two Different Multimodal Approaches to Segment 3D Ischemic Stroke Segmentation Using Brain Tumor Representation Among Multiple Center Datasets. *J. Imaging Inform. Med.* **2024**, *37*, 2375–2389. [CrossRef] [PubMed]
11. Soh, W.K.; Rajapakse, J.C. Hybrid UNet transformer architecture for ischemic stroke segmentation with MRI and CT datasets. *Front. Neurosci.* **2023**, *17*, 1298514. [CrossRef] [PubMed]
12. Ahmed, R.; Al Shehhi, A.; Werghi, N.; Seghier, M.L. Segmentation of stroke lesions using transformers-augmented MRI analysis. *Hum. Brain Mapp.* **2024**, *45*, e26803. [CrossRef] [PubMed]
13. Gui, C.; An, X.; Li, T.; Liu, S.; Ming, D. St-RegSeg: An unsupervised registration-based framework for multimodal magnetic resonance imaging stroke lesion segmentation. *Quant. Imaging Med. Surg.* **2024**, *14*, 9476–9496. [CrossRef] [PubMed]
14. Garcia-Salgado, B.P.; Almaraz-Damian, J.A.; Cervantes-Chavarria, O.; Ponomaryov, V.; Reyes-Reyes, R.; Cruz-Ramos, C.; Sadovnychiy, S. Enhanced Ischemic Stroke Lesion Segmentation in MRI Using Attention U-Net with Generalized Dice Focal Loss. *Appl. Sci.* **2024**, *14*, 8183. [CrossRef]
15. Cao, Y.; Qin, T.; Liu, Y. Exploiting Cross-modal Collaboration and Discrepancy for Semi-supervised Ischemic Stroke Lesion Segmentation from Multi-sequence MRI Images. *J. Digit. Imaging. Inform. Med.* **2026**, *39*, 2352–2364. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.