Journal of Computer Science

Research Article

Advancing Stroke Lesion Segmentation With U-Net Variants: Classical, Transfer Learning, and MRI Sequence-Specific Customized Approaches

Sonia Flora Panesar and Amit P. Ganatra

Department of Computer Science and Engineering, Faculty of Computer Science and Engineering, Parul University, Waghodia, Vadodara, India

Article history
Received: 13-03-205
Revised: 26-06-2025
Accepted: 18-09-2025

Corresponding Author:
Amit P. Ganatra
Department of Computer Science and Engineering, Faculty of Computer Science and Engineering, Parul University, Waghodia, Vadodara, India
Email: provost@paruluniversity.ac.in

Abstract: Ischemic stroke, caused by restricted cerebral blood flow, is a leading cause of death and disability worldwide. Accurate segmentation of stroke lesions in MRI is essential for timely diagnosis but remains labor-intensive when done manually. This study presents a comparative evaluation of three U-Net variants for automated ischemic stroke lesion segmentation using the ISLES 2022 dataset: (1) Classical U-Net with a standard encoder-decoder structure, (2) Transfer Learning-Enhanced U-Net with MobileNetV2 encoder pre-trained on ImageNet, and (3) A novel MRI Sequence specific Customized U-Net that employs separate modality-specific encoders for DWI,ADC and FLAIR sequences followed by fused decoding. All models were trained and evaluated using Dice Score and Dice Loss metrics. The proposed customized U-Net outperformed the other two models in a single train-validation setup, achieving a Training Dice Score of 0.8680 and a validation Dice Score of 0.8409. The architecture demonstrates robust, efficient, and accurate segmentation, addressing class imbalance and small lesion challenges. These findings highlight the potential of modality-specific architectures to enhance clinical workflows and support automated stroke diagnosis.

Keywords: Ischemic Stroke, Lesion Segmentation, UNET, Multi-Sequence MRI, Convolutional Network

Introduction

The most widespread form of brain trauma and a major contributor to disability and death worldwide is still ischemic stroke. It arises from an obstruction in cerebral blood flow, which prevents oxygen and essential nutrients from reaching brain tissue, leading to cellular death and irreversible damage. Depending on the time since the stroke onset, lesions can be categorized into acute (immediate damage to brain tissue), sub-acute (inflammatory and reparative responses occurring days to weeks post-stroke), and chronic phases (long-term structural changes like scarring or tissue atrophy) (Petzsche et al., 2022). cerebral blood flow, which prevents oxygen and essential nutrients from reaching brain tissue, leading to cellular death and irreversible damage. Depending on the time since the stroke onset, lesions can be categorized into acute (immediate damage to brain tissue), sub-acute (inflammatory and reparative responses occurring days to weeks post-stroke), and chronic phases (long-term structural changes like scarring or tissue atrophy) (Petzsche et al., 2022). Accurate assessment of affected brain regions – especially the salvageable penumbra, is critical in guiding treatments like mechanical thrombectomy (Clèrigues et al., 2019). Manual Segmentation of these lesions on MRI is labor-intensive and subject to variability especially due to their irregular shape and low contrast. Automatic segmentation offers precise structural and functional information by categorizing each voxel as a lesion or non-lesion on each slice for better decision on MRI sequences T1, T2, DWI, FLAIR etc.

Recent advancements in neuroscience image-based analysis, particularly using Computer Tomography (CT) and Magnetic Resonance Imaging (MRI), have significantly improved patient outcomes by guiding revascularization strategies. The ISLES Challenge (Ischemic Stroke Lesion Segmentation) serves as a benchmark platform for evaluating automated segmentation algorithms, offering datasets specifically

designed to advance ischemic stroke lesion segmentation techniques. The ISLES 2022 challenge dataset comprises multi-centre 3D MRI sequences including Diffusion-Weighted Imaging (DWI), Apparent Diffusion Coefficient (ADC) and Fluid-Attenuated Inversion Recovery (FLAIR), which together provide a comprehensive representation of ischemic lesions across different stages of stroke progression(Siddique et al., 2022). Various open-source tools to facilitate advanced lesion segmentation, analysis, registration, rendering and visualization of 3D and 4D biomedical images includes AMIRA, MIPAV ITKSNAP and 3DSlicer. These tools support different image formats like NIFTI and DICOM, however manual intervention is still required (Malik et al., 2024). The cross-sectional views of FLAIR 3D session from ISLES 2022 dataset shown in Figure 2 using 3D Slicer.

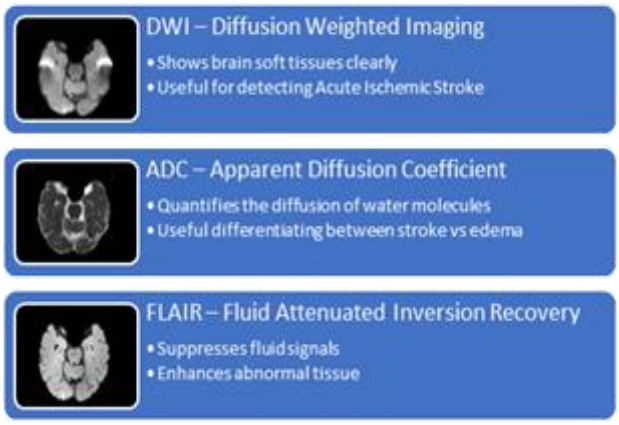

Fig. 1: MRI -Sequences in ISLES 2022 Dataset

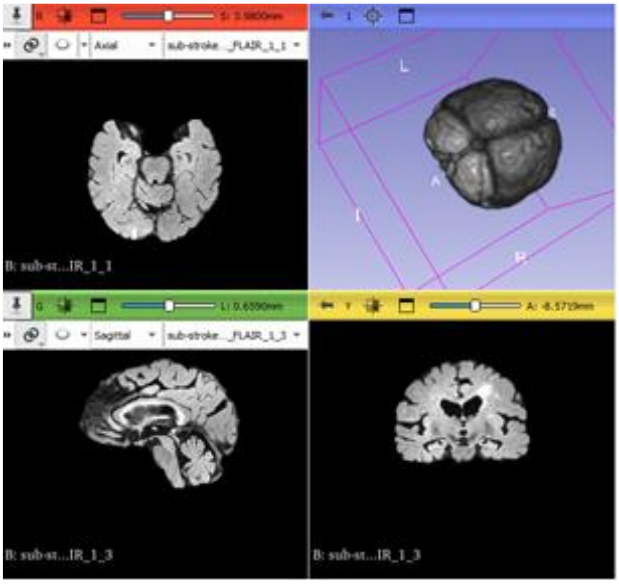

Fig. 2: FLAIR – cross-sectional views (Axial, Coronal and Sagittal) + 3D Rendering

The ISLES challenges were initiated to promote fair comparison and benchmarking of automated ischemic stroke segmentation methods by providing publicly available annotated multispectral MRI datasets. It addressed the common issue of inaccessible private datasets by offering standardized data and evaluation metrics. The 2015 challenge featured two sub-tasks: SISS (Sub-acute Ischemic Stroke Lesion Segmentation) using multi-sequence MRI, and SPES (Stroke Perfusion Estimation) focused on predicting infarct cores from CT perfusion imaging. These challenges received broad participation and have played a key role in advancing methods for ischemic stroke lesion segmentation (Maier et al., 2017). In the ISLES 2015 challenge, Maier et al. proposed a Random Forest-based segmentation method for the SISS task using carefully selected voxel-wise features from T1, T2, FLAIR, and DWI sequences. Their method incorporated spatial, intensity, and hemispheric difference features, and achieved promising results with a Dice score of 0.58 across 28 training cases. While the approach demonstrated robustness to noise and clinical variations, performance was limited by challenges such as lesion heterogeneity and under-representation of positive lesion voxels in training data (Maier et al., 2017).

U-Net introduced by Ronneberger et al. (2015) being one of the most widely used CNN based deep learning architectures for biomedical image segmentation. It consists of an encoder-decoder framework with skip connections, where the encoder captures low-resolution contextual features and the decoder progressively reconstructs high-resolution outputs. While skip connections help retain spatial details. Ronneberger et al. (2015) demonstrated strong performance on small biomedical datasets such as ISBI Cell Tracking, achieving dice score of 0.91 for segmenting cell boundaries. However standard U-Net models rely on convolution operations with small receptive fields, which limits their ability to model long-range spatial dependencies. This limitation is particularly critical for ischemic stroke segmentation, where lesions may be small, scattered and poorly contrasted against healthy tissue. The effectiveness of U-Net has been validated across domains such as retinal vessel segmentation in the DRIVE and STARE datasets (Fu et al., 2016) where an AUC of 0.95 was achieved. While highly precise, these models are sensitive to illumination and struggle with thin vessel misclassification. Similarly, U-Net-based models have been widely adopted for brain tumor segmentation in the BraTS 2015 dataset, reporting Dice scores in the range of 0.85–0.89 for tumor core and enhancing tumor regions (Menze et al., 2015). However, boundary differentiation between tumor and edema remains a challenge. In the ISLES 2018 Challenge, (Winzeck et al., 2018) benchmarked U-Net variants for multispectral MRI stroke lesion segmentation, achieving Dice scores up to 0.85. Despite strong performance, segmenting small and diffuse infarcts proved difficult, highlighting the need for more

modality specific architecture in stroke applications.

To overcome these limitations, researchers have integrated residual connections and transfer learning into U-Net architectures. Bhalerao and Thakur (2020) extended U-Net with 3D residual connections trained on BraTs 2019 dataset achieved mean dice score of 0.828. However, its performance dropped to 0.697 for enhancing tumor regions, due to class imbalance and low contrast in MRI. Shehab et al. (2021) incorporated ResNet-50 into U-Net's encoder, for faster and more accurate segmentation on BraTs2015 dataset achieved dice score as 86% with faster convergence. Yet, the approach incurred pixel-wise computation overhead in up sampling and required tuning of convolutional blocks for different modalities. More broadly, transfer learning using pretraining networks such as VGG19 or ResNet has become popular in medical imaging tasks but often struggles with domain adaptation, especially when applied to MRI or CT imaging data.

To address the challenge of acquiring high-quality annotated datasets for medical image segmentation, Antonelli et al. (2019) introduced the Medical Segmentation Decathlon (MSD), a diverse, multi-organ dataset designed to evaluate generalizable segmentation algorithms. Although MSD spurred methodological progress, it does not include ischemic stroke-specific modalities like DWI or FLAIR, which limits its relevance in stroke research. Rekik et al. (2012) provided a comprehensive review of segmentation and modeling techniques for ischemic stroke and highlighted major limitations, such as lesion swelling, reperfusion effects, and variability in lesion evolution. These biological complexities, combined with underutilized multimodal MRI data, create substantial challenges for accurate automated stroke lesion segmentation. In parallel, Zhuang (2019) proposed LadderNet, a novel multi-path U-Net variant with multiple encoder-decoder pairs designed to improve feature reuse and learning capacity. Although successful on retinal segmentation datasets (e.g., DRIVE, CHASE_DB1), LadderNet's performance and scalability for multi-sequence MRI in stroke segmentation remain underexplored.

Recent advancement in deep learning for medical image segmentation have focused on leveraging modality-specific pathways to better utilize multispectral MRI data. Dolz et al. (2019a) proposed a dense multi-path U-Net, implemented on the ISLES 2018 dataset, that processes each image modality (e.g., DWI, CBF, etc.) in separate streams and applies dense connections both within and across streams. It outperformed traditional early and late fusion U-Net approaches with higher Dice score, lower Modified Hausdorff Distance (MHD), and higher volumetric similarity (VS). Expanding this idea, Dolz et al. (2019b) also introduced HyperDenseNet, where every layer in each modality-specific path is connected to all subsequent layers in every path. This extensive connectivity promotes richer feature reuse and improved gradient flow. HyperDenseNet was evaluated on two benchmark datasets: The iSEG-2017 dataset (for 6-month infant brain tissue segmentation) and the MRBrainS dataset (for adult brain segmentation), emphasizing its strength in cross-modal feature integration. Zhang et al. (2020) presented AResU-Net, integrating attention and residual blocks which significantly improved performance with dice score of 88% on BraTs2017 dataset on 2D slices. However, its attention modules added computational overhead and were less effective in low-contrast settings. Alom et al. (2018) proposed RU-Net and R2U-Net, combining recurrent convolutional layers with residual units for enhanced feature reuse. It demonstrated superior accuracy and parameter efficiency across diverse medical image datasets including retina, skin cancer, and lung lesion. However, this architecture also introduced computational overhead due to recurrence with training time increases as recurrent depth ($t>2$) increased. Finally Wu et al. (2023a) proposed W-Net, a hybrid CNN-Transformer framework with BDM and BCM modules that achieved strong performance on ISLES 2022 and ATLAS datasets. It introduced boundary detection modules for enhanced segmentation accuracy, but was limited to 2D slices and required significant computational resources, restricting its scalability for clinical deployment.

Despite substantial advancements in deep learning-based segmentation techniques, several challenges remain. These include the variability in lesion appearance across patients, the effective integration of multi-sequence MRI data, and the demand for high segmentation accuracy despite limited training samples. Although several advanced architectures such as Dense Multipath architecture by Dolz et al. (2019a), LadderNet by Zhuang (2019), HyperDenseNet by Dolz et al. (2019b), and W-Net by Wu et al. (2023b) have been proposed for medical image segmentation, there is still ambiguity in the community regarding the optimal stage early, middle, or late for fusing multi-sequence MRI data to achieve accurate ischemic stroke segmentation. U-Net, when trained properly, remains a highly effective and robust architecture despite its relative simplicity. However, a direct and systematic comparison of U-Net variants namely classical U-Net, transfer learning-based U-Net, and a sequence-specific multi-encoder U-Net on a common benchmark like ISLES 2022 using a uniform preprocessing and followed by feature fusion at the bottleneck and shared decoding, achieving a balance between segmentation accuracy and computational efficiency. Unlike attention-heavy models, our design avoids additional attention modules to reduce GPU memory usage and inference time, making it more feasible for real-time or resource-constrained clinical settings. Most existing studies either focus on a single model or introduce novel architectures without benchmarking them against such practical and interpretable baselines. Moreover, few works analyze

training dynamics (e.g., loss/score progression), report computational efficiency, or consider deployment feasibility in resource-constrained environments. Additionally, while multimodal fusion of different imaging types such as CT and MRI is theoretically desirable, it is rare to obtain both modalities for the same patient, making such fusion strategies infeasible in most real-world clinical datasets. This creates a meaningful gap in understanding the practical trade-offs between model complexity and segmentation performance under realistic constraints for ischemic stroke lesion segmentation. Despite the widespread use of U-Net and its variants in medical image segmentation, their application to ischemic stroke segmentation particularly on the ISLES 2022 dataset, remains limited by several factors. This study aims to address these limitations by evaluating and comparing three distinct U-Net variants on the ISLES 2022 dataset, using Dice Score and Dice Loss as key performance metrics.

Methods

Our proposed Methodology involves three U-Net variants, classical U-Net, Transfer learning enhanced U-Net and Customized U-Net which will be trained on ISLES2022 dataset having three MRI sequences (DWI, ADC and Flair). Rosa et al. (2024) focuses exclusively on DWI, which is crucial for stroke diagnosis due to its sensitivity in detecting acute ischemic lesions, the lack of multi-sequence integration (e.g., FLAIR, ADC, T1, T1c) could limit the model's robustness across different imaging conditions. Shah et al. (2020) utilizes ISLES 2015 dataset, incorporating four MRI modalities (DWI, T1, T1c, FLAIR) to improve segmentation accuracy and achieved a 71% Dice Coefficient (DC) but lacks in generalizability due to limited dataset size (only 28 Patients). By concerning both factors dataset size as well as integration of different sequences, our proposed architectures are explained under different sub sections. These were implemented using the PyTorch deep learning framework due to its flexibility and dynamic computation graph support. All training and evaluation experiments were performed on Google Colab with access to an NVIDIA Tesla T4 GPU and A100 GPU, which enabled efficient GPU acceleration for handling high-dimensional MRI data. The models were trained end-to-end using the Adam optimizer with Dice Loss as the objective function. The flowchart in Figure 3 shows the key processes occurred during the implementation of each model. The pre-processing pipeline for the ISLES 2022 dataset was meticulously developed for our work to standardize MRI data across modalities and prepare it for deep learning-based segmentation tasks. The first step involved loading the MRI sequences (DWI, ADC, FLAIR) along with their respective segmentation masks (msk.nii) using the NiBabel, cv2 and numpy libraries. To ensure uniform input dimensions across all samples, the 3D MRI volumes and masks were resized to a standard voxel shape of 128×128×128 using interpolation techniques provided by the Scikit-Image library. In biomedical image pre-processing, the desired outcome requires precise localization, ensuring that each pixel is assigned a corresponding class label (Tursynova and Omarov, 2021). To facilitate computational efficiency, each 3D MRI volume was sliced along the axial plane to produce 128 2D slices per modality (DWI, ADC, and FLAIR). Corresponding segmentation masks were processed in the same way to ensure perfect alignment. All slices were resized to a uniform shape and saved as .npy files for efficient loading during training. Across 250 MRI sessions and 4 data types (3 modalities + 1 mask), this would result in a theoretical maximum of 384,000 2D slices. However, many slices particularly those at the top and bottom of the volumes were completely black (non-informative), lacking brain tissue or lesion content. These were excluded during preprocessing to eliminate irrelevant data. After filtering, the final dataset comprised 108,676 meaningful 2D slices, or approximately 27,169 slices per modality and mask. This selective approach ensured that only clinically relevant data was used, improving training efficiency and model focus.

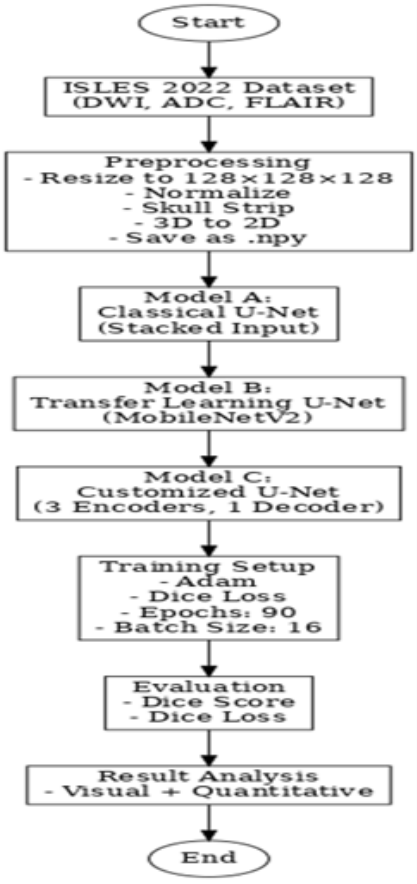

Fig. 3: Key Processes in implementing UNET -Variants during lesion segmentation

Classical U-Net

Among convolutional neural networks, U-Net stands out as one of the most significant architectures, offering high-quality medical image segmentation even with limited training data (Jiangtao et al., 2025). The classical U-Net was designed following a standard encoder-decoder architecture with skip connections. Each of the four convolutional blocks has two convolutional layers, max-pooling for down sampling, and ReLU activations. Four transposed convolutional layers for up sampling, with concatenation of features from the corresponding encoder layer (skip connections) to preserve spatial details. The layer of the outcome used a sigmoid activation to predict probabilities for the binary segmentation task (stroke lesion vs. background). The architecture of U-Net is shown in Figure 4.

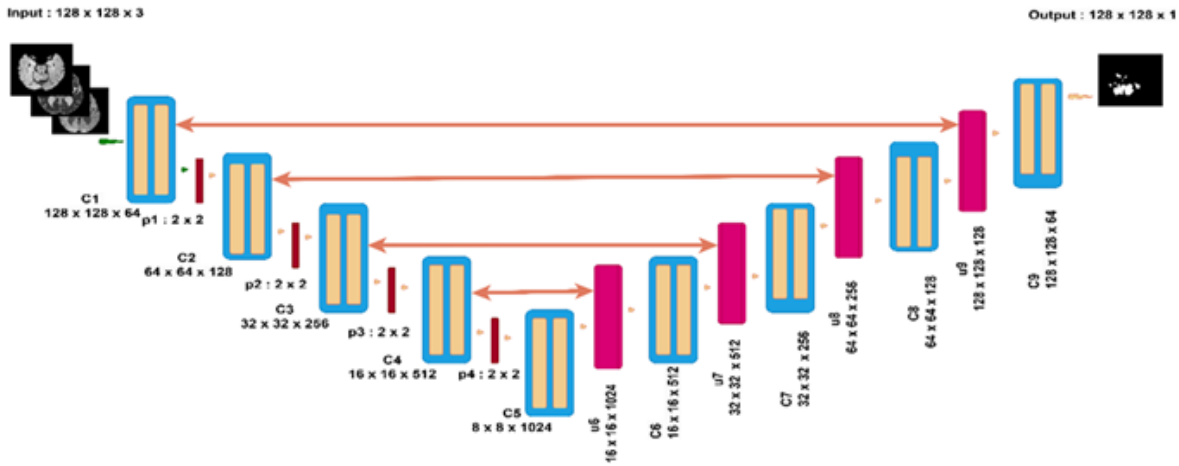

Fig. 4: Classical UNET used for stacked MRI-sequences as input and segmented mask as output

Transfer Learning Enhanced Mobile-Net U-Net

As discussed in Section 1, we use transfer learning to leverage pretrained ImageNet weights and improve feature extraction, especially in domains with limited annotated medical data. In such cases, cross domain transfer learning where models pretrained on natural images datasets are fine tuned for medical imaging tasks. It can significantly enhance feature learning and detection accuracy. In contrast, Cross-modal transfer learning applies to transferring knowledge across modalities within the same domain like pretrained model on MRI can be used to transfer its knowledge in case of Ultrasound data (Cheng and Lam, 2021). CNN-based models such as VGG16, VGG19, ResNet, and DenseNet are widely utilized as pretrained architectures in medical imaging tasks like segmentation, classification, and detection, owing to their ability to extract deep hierarchical features and improve feature propagation. In the domain of neurodegenerative diseases such as Alzheimer's, deep learning techniques have played a pivotal role in advancing early diagnosis and monitoring disease progression. For example, Polater and Sevli (2024) reported a 99.92% classification accuracy for Alzheimer's diagnosis using MobileNetV2, demonstrating its potential for robust medical diagnostic applications. Motivated by such findings, our proposed methodology, the transfer learning-enhanced U-Net integrates a MobileNetV2 encoder pretrained on the ImageNet, replacing the standard encoder in the classical U-Net architecture. MobileNetV2, introduced by Sandler et al. (2018) employs depth-wise separable convolutions and an inverted residual structure with linear bottlenecks, which makes it highly efficient for feature extraction with reduced computational complexity. These properties are particularly advantageous for data-scarce scenarios like ischemic stroke lesion segmentation. Similar encoder replacement strategies have shown promise in recent studies. Notably, Aboussaleh et al. (2023) proposed an efficient U-Net variant incorporating three parallel pretrained encoders VGG-19, ResNet50, and MobileNetV2 combined via a Bidirectional Feature Pyramid Network (BiFPN) and followed by an attention-based decoder. Their model, evaluated on the BraTS 2020 dataset, achieved Dice scores of 87.41% for whole tumor, 80.69% for core, and 70.33% for enhancing tumor, highlighting the performance gains achievable through multi-encoder transfer learning and hybrid feature fusion techniques. In our model, MobileNetV2 extracts both low-level and high-level features from the MRI sequences (FLAIR, DWI, ADC). These hierarchical features are passed to the decoder through skip connections, preserving spatial context, Blockwise configuration is shown in Table 1. The bottleneck layer consists of two

convolutional layers with 1024 filters and ReLU activations, followed by a decoder that mirrors the classical U-Net, employing transposed convolutions and skip connections. The final output layer uses sigmoid activation to generate a binary stroke lesion segmentation map. This architecture as shown in Figure 5. balances the generalization strength of pretrained CNNs with the spatial detail preservation of U-Net, making it an efficient and accurate approach for stroke lesion segmentation using limited medical data.

Customized U-NET With Separate Encoders and Single Decoder

Our primary proposed methodology based on to integrate different modalities efficiently includes customized U-Net architecture which was specifically designed to leverage the unique features of multiple sequences of MRI data (FLAIR, DWI, ADC) while preserving sequence-specific information throughout the segmentation pipeline (Wu et al., 2023a). Unlike the classical and transfer learning-enhanced U-Net approaches, this architecture employs separate encoder paths for each input sequence and combines their feature maps strategically during the decoding phase to ensure optimal fusion of information. These encoders consist of four convolutional neural blocks with max-pooling, multiple normalization The ReLU algorithm activations, and two layers of convolution each for down sampling. The separate encoders allow each sequence to undergo independent feature extraction, capturing unique spatial and textural patterns without cross-modal interference. The outputs from the deepest layers of the three encoders are concatenated into a single representation. Two convolutional layers with 1024 filters each are applied to this fused representation using ReLU activations. To reduce overfitting and improve generalization, Dropout layers (with a rate of 0.3) were added after the final encoder layers and the bottleneck stage. These layers randomly deactivate a portion of neurons during training, forcing the model to learn more robust features across folds. This fusion at the bottleneck stage allows the model to learn joint feature representations from all three modalities (Aboussaleh et al., 2023). The decoder reconstructs the segmentation map using up sampling layers and integrates modality-specific features via skip connections. The outputs from the three modalities' respective encoder blocks are merged and transferred to the following layer at each decoding stage. This hierarchical feature fusion ensures that fine-grained spatial information and high-level contextual features are effectively combined. A 1x1 convolutional layer with a sigmoid activation function is used in the ultimate output layer in order to create a binary segmentation map. The architecture is shown in Figure 6. The training process for the Classical U-Net, Transfer Learning-Enhanced U-Net, and Customized U-Net with Separate Encoders followed a structured approach, with slight variations in hyper parameters and training strategies to optimize each architecture for ischemic stroke lesion segmentation. Dice Loss was chosen as the major loss function because it effectively addresses class imbalance and tiny lesion sizes in medical images segmentation tasks, and the Adam optimizer was utilized because of its adaptive learning capabilities (Garcia-Salgado et al., 2024). To ensure fine-tuned optimization, a scheduler for learning rates was used to dynamically lower the learning rate by a factor of 0.01 if the verification loss did not improve for four successive epochs. Furthermore, training was ceased early if the validation Dice Loss did not change for ten successive epochs.

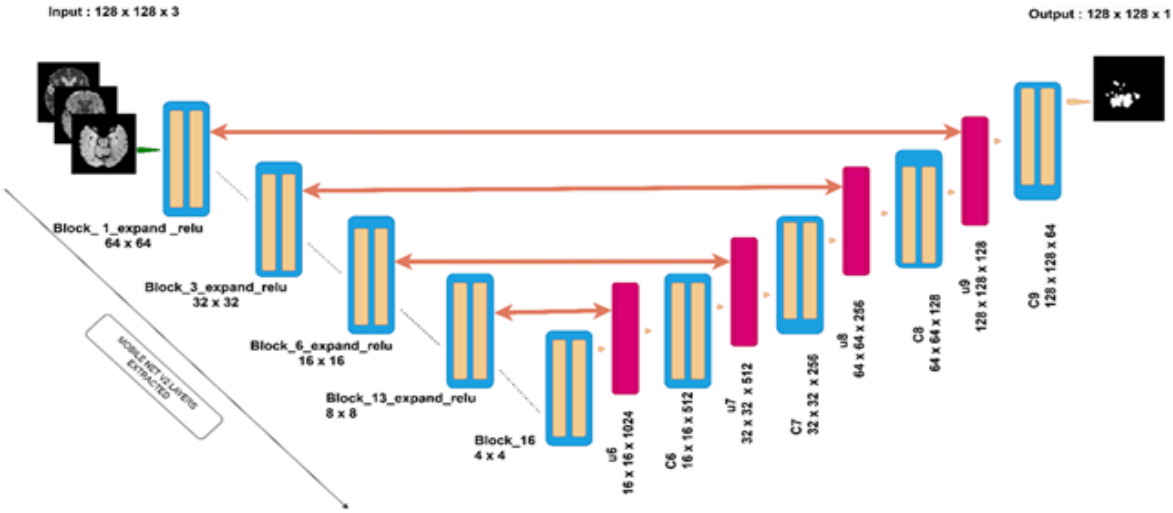

Fig. 5: MobileNET V2 as transfer learning used as Encoder in Classical UNET for stacked MRI-sequences as input and segmented mask as output

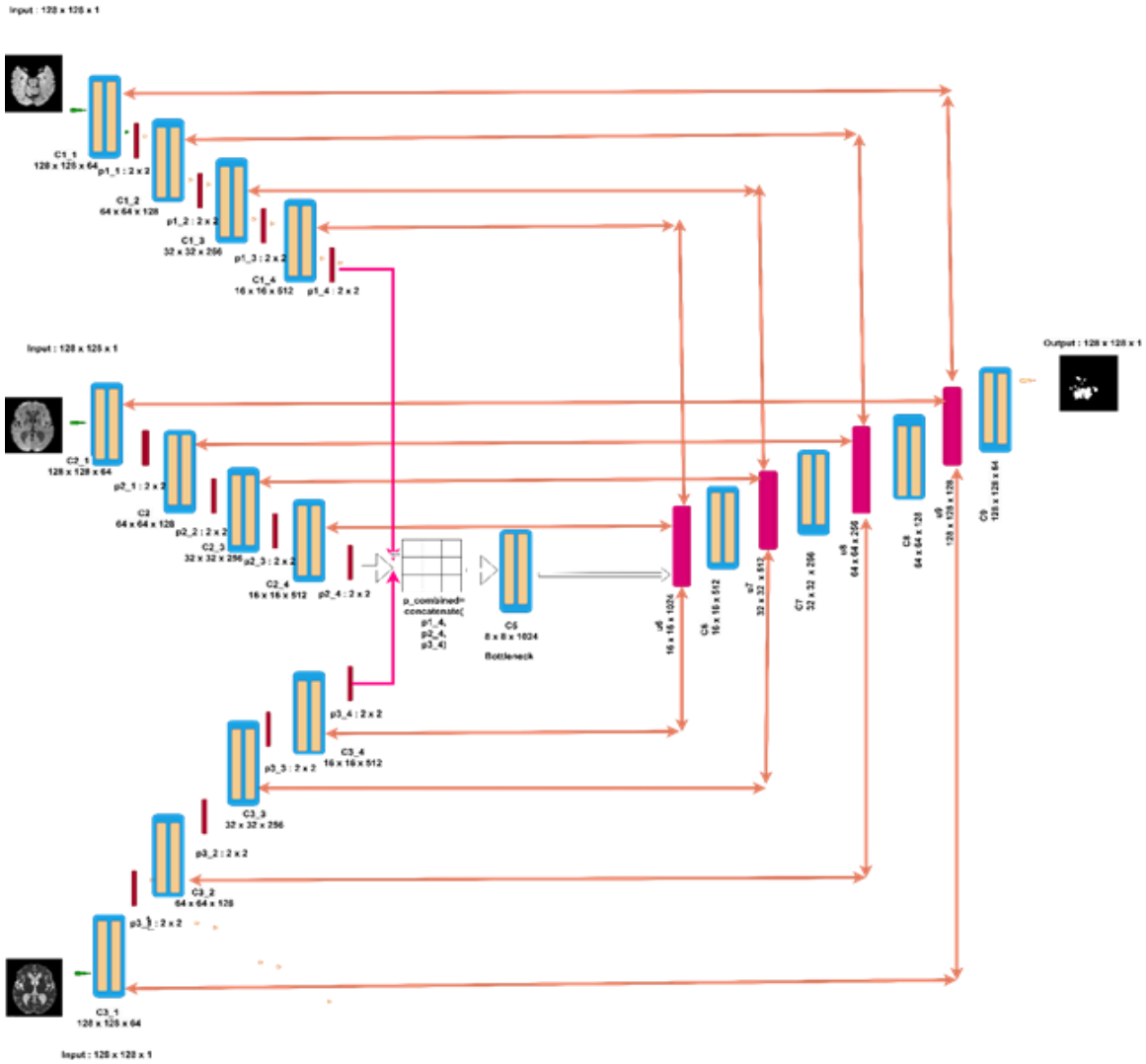

Fig. 6: Customized U-Net as separate encoder for each MRI-sequences as input and single decoder for segmented mask as output

Table 1: Block wise Configuration of MobileNet V2

Block No.	Expansion Factor	Filter	Input Size	Output Size	Stride
BN_1	1x	16	64 × 64 × 32	64 × 64 × 16	1
BN_2–3	6x	24	64 × 64 × 16	32 × 32 × 24	2
BN_4–6	6x	32	32 × 32 × 24	16 × 16 × 32	2
BN_7–10	6x	64	16 × 16 × 32	16 × 16 × 64	1
BN_11–13	6x	96	16 × 16 × 64	8 × 8 × 96	2
BN_14–16	6x	160	8 × 8 × 96	4 × 4 × 160	2
BN_17	6x	320	4 × 4 × 160	4 × 4 × 320	1

Evaluation Metrics

In line with the ISLES 2022 challenge guidelines (Petzsche et al., 2022), we acknowledge that relying solely on overlap-based metrics such as the Dice Similarity Coefficient (DSC) may not fully capture segmentation accuracy particularly for small punctiform infarcts. The ISLES organizers recommend complementary lesion-wise metrics that account for clinical interpretability, such as lesion count accuracy, lesion presence/absence, infarct volume estimation, and lesion detection F1-score. However, in this study, we focused on voxel-wise metrics including Dice Score, IoU, Precision, Recall, and Hausdorff Distance to evaluate segmentation quality. These metrics provide a strong baseline for assessing model performance and will be

extended in future work with lesion-wise evaluations to align more closely with the challenge's ranking strategy and clinical priorities.

Dice Similarity Coefficient (DSC)

The overlap between the ground truth and the projected segregation is measured by the Dice Similarity Coefficient which is shorthanded as (DSC), focusing on the balance between true positives (correctly identified lesion pixels) and both false positives (pixels incorrectly predicted as lesion) and false negatives (lesion pixels missed by the prediction). It is calculated using the formula:

$$DSC = \frac{2\ X\ TP}{2\ X\ TP + FP + FN}$$

The Dice coefficient balances precision (how many predicted lesions are correct) and recall (how many actual lesions are correctly identified). It ranges from 0 for no overlap to 1 for perfect overlap, making it an ideal metric for assessing segmentation performance, particularly in medical images where lesions are sparse compared to the background (Sudre et al., 2017). Dice score near to 1 indicates that the predicted mask closely matches the ground truth, with minimal false positives and false negatives, ensuring accurate lesion identification.

Dice Loss

The Dice Loss, being the complement of the loss function termed the Dice Similarity Coefficient (DSC) was created expressly to reduce the imbalance between the ground truth and the anticipated segmentation, particularly when the foreground (such as lesion zones) makes just a tiny percentage of the the image (Sudre et al., 2017). Prior work has shown that optimizing cross-entropy loss while evaluating with Dice or IoU metrics leads to a mismatch, reducing segmentation performance (Bertels et al., 2019). Surrogate losses like soft-Dice or soft-Jaccard are preferred to better align the training objective with evaluation metrics.The Dice Loss is defined as:

$$Dice\ Loss = 1 - DSC$$

In terms of TP, FP and FN, this becomes:

$$Dice\ Loss = 1 - \frac{2\ X\ TP}{2\ X\ TP + FP + FN}$$

Minimizing Dice loss ensures the model learns to optimize the overlap for predictions and the ground truth while reducing the impact of incorrect predictions (FP) and missed detections (FN), leading to better segmentation performance:

Intersection Over Union (IoU)

IoU, also known as the Jaccard Index, measures the overlap between the predicted and ground truth masks relative to their union:

$$IoU = \frac{TP}{TP + FP + FN}$$

While Dice and IoU are related, tends to penalize mismatches more heavily and complements the Dice Score.

Precision

Precision, quantifies how many of the predicted lesion pixels are correct:

$$Precision = \frac{TP}{TP + FP}$$

Hausdorff Distance (HD)

HD measures the maximum distance between the predicted(P) and ground truth (A) contours. It reflects boundary accuracy and is sensitive to outliers, making it important for evaluating segmentation of small or irregularly shaped lesions (Garcia-Salgado et al., 2024).

$$HD(P,A) = max(\max_{p\in P}\min_{a\in A} d(p,a), \max_{a\in A}\min_{p\in P} d(a,p)$$

Results and Discussion

This section presents a comprehensive evaluation of the three implemented U-Net architectures using the ISLES 2022 dataset. The models were compared based on segmentation accuracy,convergence behaviour, computational efficiency and generalization performance across cross-validation folds. Each model was assessed using key segmentation metrics: Train Dice Score, Validation Dice Score, Train Loss, and Validation Loss, are shown below in the Table 2.

The classical U-Net recorded the lowest performace, with a validation Dice Score of 0.5966, indicating limited generalization capability. The Transfer Learning-Enhanced U-Net improved significantly, achieving a validation Dice Score of 0.7844, benefiting from pretrained MobileNetV2 features. The Customized U-Net with Separate Encoders outperformed both, with the highest validation Dice Score of 0.8409 and lowest validation loss, highlighting the advantage of modality-specific feature extraction and fusion. The feasibility of deploying these architectures in clinical environments was assessed through model size, inference time per 2D slice, and GPU memory usage, summarized in Table 3. The Transfer Learning U-Net offered the most lightweight implementation, with the lowest inference

time and memory usage. In contrast, the Customized U-Net incurred higher computational cost due to its multi-branch encoder structure but provided superior segmentation accuracy. All models remained within acceptable limits for near-real-time clinical use (<1.2 seconds per slice).

Table 2: Dice Score and Dice Loss among three U-Net Variants

Model Meic	Classical U-Net	Transfer Learning U-Net	Customized U-Net
Training Dice Score	0.6507	0.8243	0.8680
Validation Dice Score	0.5966	0.7844	0.8409
Training Loss	0.4409	0.237	0.1583
Validation Loss	0.4793	0.266	0.1772

Table 3: Computational performance comparison of the three U-Net variants

Metric	Classical U-Net	Transfer Learning U-Net (MobileNetV2 Encoder)	Customized U-Net (Separate Modality Encoders)
Parameter Count	~7.8 million	~3.6 million	~10.4 million
Average Inference Time (per 2D slice)	0.72 sec	0.58 sec (fastest)	1.12 sec (slowest)
Peak GPU Memory Usage	~5.4 GB	~4.8 GB(lowest)	~6.7 GB

As shown in Figure 7, the Customized U-Net with separate modality-specific encoders consistently outperformed both the Classical and Transfer Learning-based U-Nets across all four evaluation metrics. However, the classical U-Net lagged behind both in accuracy and convergence behavior, confirming the limitations of a single-encoder structure when handling multi-modal MRI data.

The plot in Figure 8 compares the validation dice score trends across 25 epochs for U-Net variants. The Customized U-Net consistently achieves the highest validation Dice score, demonstrating superior segmentation accuracy and convergence behavior compared to the other models. The Transfer Learning U-Net shows stable improvement, while the Classical U-Net lags behind in both convergence and final performance.

To ensure robust evaluation, the customized U-Net with separate encoders was further validated using 5-fold cross-validation. This approach assesses the model's generalizability across varying data splits and reduces the risk of overfitting. The evaluation metrics included Validation Dice Score, Intersection over Union (IoU), Precision, Recall, Hausdorff Distance, and Inference Time per slice, as summarized in Table 4.

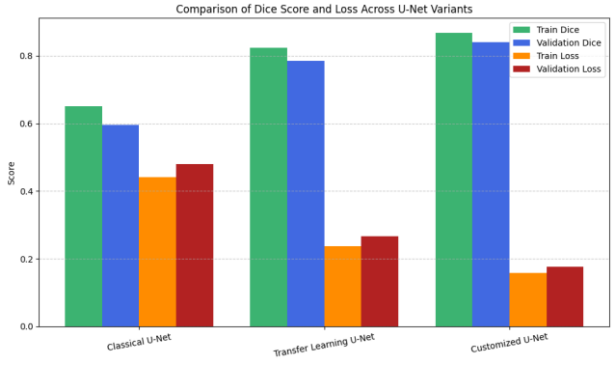

Fig. 7: Comparison of Dice Score and Loss Metrics across U-Net variants

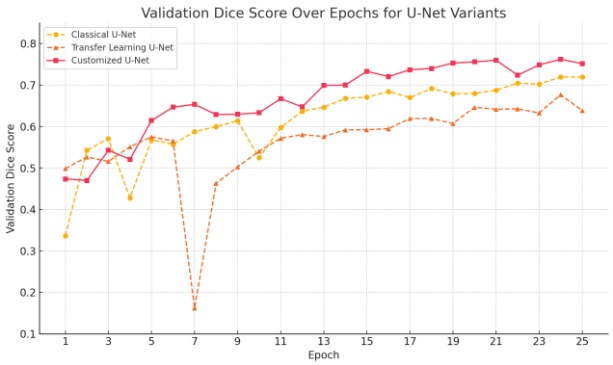

Fig. 8: Validation Dice Score Over Epochs for U-Net Variants

Table 4: Cross Validation Metrics recorded for Customized U-Net

Metric	Mean	Standard Deviation
Validation Dice Score	0.7465	0.0748
IoU	0.5846	0.0875
Precision	0.8274	0.0511
Recall	0.682	0.095
Hausdorff	24.84	2.130
Inference Time(sec)	0.014	0.002

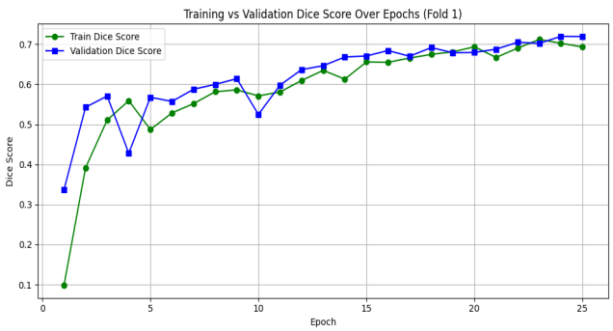

Fig. 9: Training vs. Validation Dice Score Over Epochs (Fold1)

The cross-validation metrics confirm the model's strong generalization capibility and competitive inference time, affirming its applicability in clinical scenarios for stroke lesion segmentation. Fig. 9 depicts the training and validation dice score progression across 25 epochs for the customized U-Net model during Fold 1. The training curve shows steady improvement, while the validation curve closely follows it, indicating stable generalization without overfitting.

The model achieves the peal validation dice score of 0.7197 at epoch 24. To account for minor fluctuation, the average validation dice score over the last three epochs is approximately 0.7138, refleting consistent segmentation performance as the model converges.

Qualitative Results and Discussion

The qualitative evaluation highlights the visual segmentation performance of the three U-Net variants: Classical U-Net, Transfer Learning-Enhanced U-Net, and Customized U-Net with Separate Encoders on different inputs. Visual comparisons of segmentation masks on ischemic stroke lesions are shown in Figure 10. These reveal observations as the Classical U-Net struggled with accurately delineating lesion boundaries, particularly in cases with small and irregular lesion shapes.

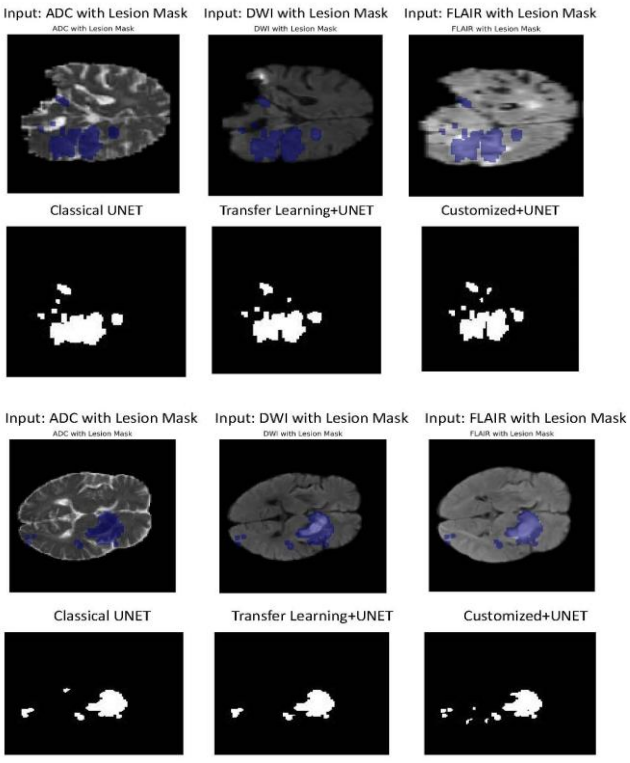

Fig. 10: Qualitative Segmentation results for two representative cases using three UNet variants

Observed issues include over-segmentation (predicting lesions larger than the actual size) and under-segmentation (missing portions of the lesion), especially in complex cases with low contrast. By leveraging the pretrained MobileNetV2 encoder, the Transfer Learning-Enhanced U-Net achieved notable improvements in segmentation quality. The model exhibited better spatial alignment of predicted masks with ground truth lesions, reducing the incidence of both false positives and false negatives. However, it occasionally struggled with fine-grained lesion boundaries, particularly for highly irregular regions. The Customized U-Net with Separate Encoders consistently delivered the most accurate segmentation masks across all cases. It excelled in preserving small and irregular lesion structures, demonstrating superior spatial accuracy , better boundary delineation and strong generalization. This performance can be attributed to its ability to extract MRI sequence-specific features from DWI, ADC, and FLAIR inputs and effectively fuse these features in the decoder.

Conclusion

This study presented a comprehensive comparative analysis of three U-Net variants Classical U-Net, Transfer Learning-Enhanced U-Net, and a Customized U-Net with Modality-Specific Encoders for ischemic stroke lesion segmentation using the ISLES 2022 dataset, which includes DWI, ADC, and FLAIR MRI sequences. The results clearly demonstrate that incorporating pretrained encoders and modality-specific processing significantly improves segmentation accuracy over the classical U-Net. The Transfer Learning-Enhanced U-Net, utilizing a MobileNetV2 encoder pretrained on ImageNet, exhibited strong generalization capabilities and reduced computational overhead. However, the highest performance was achieved by the Customized U-Net, which features separate encoder branches for each MRI modality. This architecture preserved sequence-specific features and enabled effective fusion during decoding, resulting in a validation Dice Score of 0.8409 and stable convergence across training folds. The use of Dice Loss further addressed class imbalance, improving the model's ability to segment small and irregular stroke lesions. Despite these advancements, challenges remain particularly in segmenting highly variable and subtle lesion patterns and managing the increased computational cost of multi-branch models. Future work will focus on incorporating Swin Transformer blocks into the decoder to better capture global contextual features and boundary precision, Implementing ensemble learning strategies to combine model strengths. Ultimately, this research highlights the critical role of modality-aware deep learning architectures in advancing stroke lesion segmentation and paves the way for more reliable,

generalizable, and clinically applicable tools.

Acknowledgment

I would like to express my sincere gratitude to Parul University, Vadodara for providing the necessary resources, facilities, and support that contributed to the successful completion of this research work. The encouragement and guidance from the faculty and research staff have been invaluable in shaping the study.

Author's Contributions

Sonia Flora Panesar: Writing original draft, review and editing, artwork.

Amit P. Ganatra: Conceptualization, visualization, project administration.

Ethics

The authors declare that there are no conflicts of interest related to this research work. No financial, personal, or professional affiliations influenced the study's design, data collection, analysis, or interpretation of results.

The MRI data used in this study were obtained from the publicly available ISLES 2022 challenge dataset (https://www.isles-challenge.org/ISLES2022/). All data were anonymized by the original providers, and ethical approvals and patient consents were obtained by the institutions involved in data collection, as stated by the ISLES organizers. This study complies with the data usage agreement set forth by the ISLES challenge and no additional ethical approval was required for secondary analysis.

References

Aboussaleh, I., Riffi, J., Fazazy, K. E., Mahraz, M. A., & Tairi, H. (2023). Efficient U-Net Architecture with Multiple Encoders and Attention Mechanism Decoders for Brain Tumor Segmentation. *Diagnostics*, *13*(5), 872. https://doi.org/10.3390/diagnostics13050872

Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B. A., Litjens, G., Menze, B., Ronneberger, O., Summers, R. M., van Ginneken, B., Bilello, M., Bilic, P., Christ, P. F., Do, R. K. G., Gollub, M. J., Heckers, S. H., Huisman, H., Jarnagin, W. R., Cardoso, M. J. (2019). A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *Computer Vision and Pattern Recognition*. https://doi.org/https://doi.org/10.48550/arXiv.1902.09063

Alom, M. Z., Hasan, M., Taha, T. M., & Asari, V. K. (2018). Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation. *Computer Vision and Pattern Recognition*, *6*(1), 1–12. https://doi.org/https://doi.org/10.48550/arXiv.1802.06955

Bhalerao, M., & Thakur, S. (2020). Brain Tumor Segmentation Based on 3D Residual U-Net. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, 218–225. https://doi.org/10.1007/978-3-030-46643-5_21

Bertels, J., Eelbode, T., Berman, M., Vandermeulen, D., Maes, F., Bisschops, R., & Blaschko, M. B. (2019). Optimizing the Dice Score and Jaccard Index for Medical Image Segmentation: Theory and Practice. *Proceeding of the Medical Image Computing and Computer Assisted Intervention*, 92–100. https://doi.org/https://doi.org/10.1007/978-3-030-32245-8_11

Cheng, D., & Lam, E. Y. (2021). Transfer Learning U-Net Deep Learning for Lung Ultrasound Segmentation. *Image and Video Processing*. https://doi.org/https://doi.org/10.48550/arXiv.2110.02196

Clèrigues, A., Valverde, S., Bernal, J., Freixenet, J., Oliver, A., & Lladó, X. (2020). Acute and sub-acute stroke lesion segmentation from multimodal MRI. *Computer Methods and Programs in Biomedicine*, *194*, 105521. https://doi.org/10.1016/j.cmpb.2020.105521

Dolz, J., Ben Ayed, I., & Desrosiers, C. (2019a). Dense Multi-path U-Net for Ischemic Stroke Lesion Segmentation in Multiple Image Modalities. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, *11383*, 271–282. https://doi.org/10.1007/978-3-030-11723-8_27

Dolz, J., Gopinath, K., Yuan, J., Lombaert, H., Desrosiers, C., & Ben Ayed, I. (2019b). HyperDense-Net: A Hyper-Densely Connected CNN for Multi-Modal Image Segmentation. *IEEE Transactions on Medical Imaging*, *38*(5), 1116–1126. https://doi.org/10.1109/tmi.2018.2878669

Fu, H., Xu, Y., Lin, S., Kee Wong, D. W., & Liu, J. (2016). DeepVessel: Retinal Vessel Segmentation via Deep Learning and Conditional Random Field. *Medical Image Computing and Computer-Assisted Intervention*, *9901*, 132–139. https://doi.org/10.1007/978-3-319-46723-8_16

Garcia-Salgado, B. P., Almaraz-Damian, J. A., Cervantes-Chavarria, O., Ponomaryov, V., Reyes-Reyes, R., Cruz-Ramos, C., & Sadovnychiy, S. (2024). Enhanced Ischemic Stroke Lesion Segmentation in MRI Using Attention U-Net with Generalized Dice Focal Loss. *Applied Sciences*, *14*(18), 8183. https://doi.org/10.3390/app14188183

Jiangtao, W., Ruhaiyem, N. I. R., & Panpan, F. (2025). A Comprehensive Review of U-Net and Its Variants: Advances and Applications in Medical Image Segmentation. *IET Image Processing*, *19*(1), e70019. https://doi.org/10.1049/ipr2.70019

Maier, O., Menze, B. H., von der Gablentz, J., Häni, L., Heinrich, M. P., Liebrand, M., Winzeck, S., Basit, A., Bentley, P., Chen, L., Christiaens, D., Dutil, F., Egger, K., Feng, C., Glocker, B., Götz, M., Haeck, T., Halme, H.-L., Havaei, M., … Reyes, M. (2017). ISLES 2015 - A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. *Medical Image Analysis*, *35*, 250–269. https://doi.org/10.1016/j.media.2016.07.009

Malik, M., Chong, B., Fernandez, J., Shim, V., Kasabov, N. K., & Wang, A. (2024). Stroke Lesion Segmentation and Deep Learning: A Comprehensive Review. *Bioengineering*, *11*(1), 86. https://doi.org/10.3390/bioengineering11010086

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.-A., Arbel, T., Avants, B. B., Ayache, N., Buendia, P., Collins, D. L., Cordier, N., … Van Leemput, K. (2015). The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, *34*(10), 1993–2024. https://doi.org/10.1109/tmi.2014.2377694

Petzsche, M. R. H., Rosa, E. de la, Hanning, U., Wiest, R., Pinilla, W. E. V., Reyes, M., Meyer, M. I., Liew, S.-L., Kofler, F., Ezhov, I., Robben, D., Hutton, A., Friedrich, T., Zarth, T., Bürkle, J., Baran, T. A., Menze, B., Broocks, G., Meyer, L., Kirschke, J. S. (2022). ISLES 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset. *Scientific Data*, *9*(1), 762. https://doi.org/10.1038/s41597-022-01875-5

Rosa, E. de la, Reyes, M., Liew, S.-L., Hutton, A., Wiest, R., Kaesmacher, J., Hanning, U., Hakim, A., Zubal, R., Valenzuela, W., Robben, D., Sima, D. M., Anania, V., Brys, A., Meakin, J. A., Mickan, A., Broocks, G., Heitkamp, C., Gao, S., … Wiestler, B. (2024). A Robust Ensemble Algorithm for Ischemic Stroke Lesion Segmentation: Generalizability and Clinical Utility Beyond the ISLES Challenge. *Nature Communications*, *16*(1), 7357. https://doi.org/https://doi.org/10.1038/s41467-025-62373-x

Rekik, I., Allassonnière, S., Carpenter, T. K., & Wardlaw, J. M. (2012). Medical image analysis methods in MR/CT-imaged acute-subacute ischemic stroke lesion: Segmentation, prediction and insights into dynamic evolution simulation models. A critical appraisal. *NeuroImage: Clinical*, *1*(1), 164–178. https://doi.org/10.1016/j.nicl.2012.10.003

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention*, 234–241.

Siddique, M. M. R., Yang, D., He, Y., Xu, D., & Myronenko, A. (2022). Automated ischemic stroke lesion segmentation from 3D MRI. *Image and Video Processing*, *1*(1). https://doi.org/https://doi.org/10.48550/arXiv.2209.09546

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. *2018 IEEE/CVF Conference*, 4510–4520. https://doi.org/10.1109/cvpr.2018.00474

Shah, P. M., Khan, H., Shafi, U., Islam, S. ul, Raza, M., Son, T. T., & Le-Minh, H. (2020). 2D-CNN Based Segmentation of Ischemic Stroke Lesions in MRI Scans. *Advances in Computational Collective Intelligence*, *1287*, 276–286. https://doi.org/10.1007/978-3-030-63119-2_23

Shehab, L. H., Fahmy, O. M., Gasser, S. M., & El-Mahallawy, M. S. (2021). An efficient brain tumor image segmentation based on deep residual networks (ResNets). *Journal of King Saud University Engineering Sciences*, *33*(6), 404–412. https://doi.org/10.1016/j.jksues.2020.06.001

Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., & Jorge Cardoso, M. (2017). Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, *10553*, 240–248. https://doi.org/10.1007/978-3-319-67558-9_28

Tursynova, A., & Omarov, B. (2021). 3D U-Net for brain stroke lesion segmentation on ISLES 2018 dataset. *International Conference on Electronics Computer and Computation (ICECCO)*, 1–4. https://doi.org/10.1109/icecco53203.2021.9663825

Winzeck, S., Hakim, A., McKinley, R., Pinto, J. A. A. D. S. R., Alves, V., Silva, C., Pisov, M., Krivov, E., Belyaev, M., Monteiro, M., Oliveira, A., Choi, Y., Paik, M. C., Kwon, Y., Lee, H., Kim, B. J., Won, J.-H., Islam, M., Ren, H., … Reyes, M. (2018). ISLES 2016 and 2017-Benchmarking Ischemic Stroke Lesion Outcome Prediction Based on Multispectral MRI. *Frontiers in Neurology*, *9*. https://doi.org/10.3389/fneur.2018.00679

Wu, B., Zhang, F., Xu, L., Shen, S., Shao, P., Sun, M., Liu, P., Yao, P., & Xu, R. X. (2023a). Modality preserving U-Net for segmentation of multimodal medical images. *Quantitative Imaging in Medicine and Surgery*, *13*(8), 5242–5257. https://doi.org/10.21037/qims-22-1367

Wu, Z., Zhang, X., Li, F., Wang, S., Huang, L., & Li, J. (2023b). W-Net: A boundary-enhanced segmentation network for stroke lesions. *Expert Systems with Applications*, *230*, 120637. https://doi.org/10.1016/j.eswa.2023.120637

Zhang, J., Lv, X., Zhang, H., & Liu, B. (2020). AResU-Net: Attention Residual U-Net for Brain Tumor Segmentation. *Symmetry*, *12*(5), 721. https://doi.org/10.3390/sym12050721

Zhuang, J. (2019). LadderNet: Multi-path networks based on U-Net for medical image segmentation. *Computer Vision and Pattern Recognition*. https://doi.org/https://doi.org/10.48550/arXiv.1810.07810