Journal of Computing Theories and Applications ISSN:3024-9104

Research Article

An Enhanced UNet++ with InceptionNeXt Blocks and Feature-Scale Channel Attention for Ischemic Stroke Lesion Segmentation

Muhammad Hilmy Naufal *, Wiharto*, and Herdito Ibnu Dewangkoro

Faculty of Information Technology and Data Science, Sebelas Maret University, Surakarta, 57126, Jawa Tengah, Indonesia;
e-mail :muhhilmynaufal@student.uns.ac.id; wiharto@staff.uns.ac.id; herditoibnu@staff.uns.ac.id
* Corresponding Author : Muhammad Hilmy Naufal, Wiharto

Abstract: Ischemic stroke lesion segmentation from Magnetic Resonance Imaging (MRI) remains a challenging task due to the small lesion size, irregular morphology, low contrast with surrounding tissue, and severe class imbalance. To address these challenges, this study proposes an enhanced UNet++ architecture that integrates Modified InceptionNeXt Blocks and Feature-Scale Channel Attention (FSCA) to improve multi-scale feature extraction and adaptive feature fusion. Hyperparameter tuning was performed by optimizing the number of initial filters, network depth, and loss function, followed by an ablation study to evaluate the contribution of each architectural component. The proposed model was implemented using the TensorFlow framework and evaluated on the ISLES 2022 dataset. To enhance lesion visibility, Diffusion-Weighted Imaging (DWI) was processed using Contrast Limited Adaptive Histogram Equalization (CLAHE) to generate enhanced DWI (eDWI). Three input configurations, namely DWI, DWI+ADC, and DWI+ADC+eDWI, were investigated through channel concatenation. Performance was assessed using Dice Score, Intersection over Union (IoU), Precision, and Recall. Experimental results demonstrate that the proposed UNet++ model with Modified InceptionNeXt Blocks and FSCA achieves the best performance using the DWI+ADC+eDWI configuration, obtaining a Dice Score of 0.8742, IoU of 0.8464, Precision of 0.9440, and Recall of 0.8906. Furthermore, the segmentation results exhibit high agreement with the ground truth, indicating that the proposed architecture effectively improves ischemic stroke lesion segmentation while maintaining computational efficiency.

Keywords: Computer-Aided Diagnosis; Deep Learning; Feature-Scale Channel Attention; InceptionNeXt; Ischemic Stroke; Medical Image Segmentation; Multi-Sequence MRI; UNet++.

Received: June, 9th 2026
Revised: July, 20th 2026
Accepted: July, 21st 2026
Published: Augsust, 12th 2026

1. Introduction

Ischemic stroke remains one of the leading causes of death and long-term disability worldwide, posing a major challenge to modern healthcare systems [1]. This condition occurs when blood flow to brain tissue is interrupted, resulting in neuronal cell death that may lead to permanent neurological impairment or death [2]. The effectiveness of clinical treatment largely depends on the timely and accurate identification of lesion location and extent [3]. Among various neuroimaging modalities, Magnetic Resonance Imaging (MRI), particularly Diffusion-Weighted Imaging (DWI) and the Apparent Diffusion Coefficient (ADC), has become the clinical standard because of its high sensitivity to diffusion changes occurring during the acute stage of ischemic stroke [4]. Despite these advantages, accurate lesion segmentation remains challenging because ischemic stroke lesions are typically small, exhibit irregular morphology, and often have low contrast relative to surrounding healthy tissue [5].

The intrinsic characteristics of ischemic stroke lesions further complicate automated segmentation. Lesions commonly occupy only 0.1–5% of the total brain volume, exhibit irregular boundaries caused by perilesional edema, present low contrast against adjacent healthy tissue, and suffer from severe class imbalance in which lesion pixels are substantially outnumbered by background pixels. Moreover, lesion morphology varies considerably across patients and different stages of stroke progression [6]. Traditional image segmentation techniques, including thresholding, region growing, and watershed segmentation, are highly sensitive to intensity variations, noise, and poor contrast, frequently failing to delineate small lesions or lesions located in regions with textures similar to normal brain tissue [7]. These limitations have motivated the adoption of deep learning-based approaches to improve segmentation accuracy, robustness, and consistency.

Recent advances in deep learning have significantly improved ischemic stroke lesion segmentation. Convolutional neural network (CNN) architectures, including ResNet, U-Net, and DenseNet, have demonstrated promising performance by learning discriminative lesion representations directly from imaging data [8]. More recent studies have further extended CNN-based methods for cerebrovascular analysis, such as ischemic stroke classification through pixel brightness transformation and adversarial training strategies [9]. Nevertheless, conventional CNN-based architectures rely primarily on local receptive fields, which limits their ability to preserve fine structural details and capture broader contextual information required for accurate pixel-level segmentation, particularly for small lesions [10]. Although ensemble learning and Vision Transformer (ViT)-based models have achieved encouraging results in multimodal neuroimaging applications, they mainly focus on classification tasks or require substantially larger datasets and computational resources, limiting their practicality in resource-constrained clinical environments [11], [12].

Among existing segmentation architectures, U-Net has become one of the most widely adopted models because of its encoder-decoder structure with skip connections, enabling effective learning even from relatively small medical datasets [13]. However, U-Net still suffers from semantic discrepancies between encoder and decoder feature representations, which can reduce segmentation accuracy for small or low-contrast lesions. UNet++ addresses this limitation by redesigning the skip connections through nested dense pathways that reduce the semantic gap between encoder and decoder features. Nevertheless, UNet++ continues to rely on conventional convolution blocks, which may limit its capability to capture rich multi-scale contextual information while simultaneously suppressing irrelevant background responses.

To overcome these limitations, this study integrates several complementary components into the UNet++ architecture. The Feature-Scale Channel Attention (FSCA) module is employed to adaptively emphasize lesion-relevant features while suppressing irrelevant background activations, thereby addressing the challenges posed by small lesions and severe class imbalance [14]. Meanwhile, the Modified InceptionNeXt Block enhances multi-scale feature extraction through parallel convolutional branches (1×1, 3×3, 5×5) with different receptive fields, enabling more effective representation of lesions with varying sizes and irregular morphologies [15]. To further mitigate the class imbalance problem during optimization, the proposed framework adopts the Tversky Focal Loss, which provides a better balance between precision and recall. Furthermore, multi-sequence MRI information is exploited by combining DWI and ADC through channel concatenation, allowing the network to utilize complementary anatomical and diffusion information, where DWI highlights acute hyperintense lesions and ADC provides quantitative diffusion restriction information [16].

Although DWI is highly sensitive to acute ischemic lesions, raw DWI images often exhibit limited contrast between lesions and surrounding tissue, particularly for small or subacute infarctions. Therefore, Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied to enhance DWI images, producing enhanced DWI (eDWI) with improved lesion visibility [17]. The enhanced DWI is subsequently concatenated with the original DWI and ADC images to construct a richer multi-sequence representation that preserves both the original diffusion characteristics and enhanced contrast information. Based on this design, we hypothesize that integrating the FSCA module and Modified InceptionNeXt Blocks into UNet++, together with multi-sequence MRI input and Tversky Focal Loss, can improve ischemic stroke lesion segmentation by simultaneously enhancing multi-scale feature representation, adaptive feature selection, and learning robustness under severe class-imbalanced conditions [18].

Motivated by these considerations, this study proposes an enhanced UNet++ architecture that integrates Modified InceptionNeXt Blocks and the FSCA module for ischemic stroke lesion segmentation. The proposed framework combines multi-sequence DWI, ADC, and enhanced DWI inputs through channel concatenation to exploit complementary imaging information while improving feature representation and feature fusion. Through this integrated design, the proposed model aims to achieve more accurate, robust, and computationally efficient ischemic stroke lesion segmentation, particularly for lesions that are small, faint, or exhibit low-intensity contrast.

2. Related Work

2.1. Challenges in Ischemic Stroke Segmentation

Ischemic stroke lesions on MRI exhibit several characteristics that make automated segmentation particularly challenging. These lesions are typically small, occupying only 0.1–5% of the total brain volume, have irregular boundaries caused by perilesional edema, display low contrast against surrounding healthy tissue, and suffer from severe class imbalance, where lesion pixels are substantially outnumbered by background pixels [5, 6]. Such class imbalance poses a significant challenge for deep learning models because conventional loss functions often fail to adequately penalize misclassified minority-class pixels, resulting in poor detection of small lesions [19]. In addition, traditional image segmentation approaches, including thresholding, region growing, and watershed segmentation, are highly sensitive to intensity variations and noise, frequently failing to delineate small lesions or lesions with textures similar to normal brain tissue [20]. These limitations have motivated the widespread adoption of deep learning-based methods to achieve more accurate, robust, and consistent ischemic stroke lesion segmentation [21].

2.2. UNet++ in Medical Imaging

The introduction of U-Net by Ronneberger et al. [13] marked a significant milestone in medical image segmentation. Its encoder-decoder architecture with skip connections effectively preserves spatial information while learning hierarchical contextual features from medical images. Nevertheless, the original U-Net still encounters difficulties in capturing fine structural details and reducing the semantic discrepancy between encoder and decoder representations, particularly when segmenting small or subtle lesions.

To overcome these limitations, Zhou et al. [7] proposed UNet++, which introduces nested and dense skip pathways to progressively reduce the semantic gap while improving multi-scale feature fusion. This redesign substantially enhances feature propagation and segmentation accuracy compared with the original U-Net. However, despite these improvements, UNet++ continues to rely on conventional convolution blocks, which are less effective in capturing multi-scale variations in lesion appearance. This limitation is particularly relevant for ischemic stroke lesions, whose size and morphology can vary considerably across patients and different stages of stroke progression.

2.3. Attention and Multi-Scale Feature Learning

Attention mechanisms have been widely adopted to improve feature representation by enabling neural networks to focus on informative regions while suppressing irrelevant background responses. Woo et al. [22] introduced the Convolutional Block Attention Module (CBAM), which combines channel and spatial attention to refine feature maps effectively. Building on this concept, Tagnamas et al. [14] proposed Spatial-aware Channel Attention (SCA) and Feature-Scale Channel Attention (FSCA), demonstrating that attention mechanisms can improve encoder-decoder feature interaction while maintaining parameter efficiency. Similar attention-based strategies have also been successfully integrated into U-Net-based architectures for various medical image segmentation tasks, further demonstrating their effectiveness in improving segmentation performance [23].

In parallel, multi-scale feature extraction has become an important strategy for addressing the large variability in lesion size and morphology. Yu et al. [24] introduced the InceptionNeXt block, which combines the multi-branch design of the Inception architecture with the computational efficiency of ConvNeXt through grouped depthwise convolutions. By employing parallel convolutional kernels with different receptive fields, InceptionNeXt effectively captures lesion patterns at multiple spatial scales while maintaining computational efficiency. This characteristic is particularly advantageous for medical image segmentation, where annotated datasets are often limited and achieving a balance between segmentation performance and computational cost remains an important consideration [25]. Overall, attention mechanisms and multi-scale feature extraction have become complementary strategies for improving medical image segmentation, particularly in applications involving complex lesion morphology and substantial scale variation.

2.4. Leveraging Multi-Sequence MRI Data

Beyond architectural innovations, leveraging complementary MRI modalities has become an effective strategy for improving ischemic stroke lesion segmentation. Different MRI sequences provide complementary information, allowing deep learning models to better characterize lesion appearance and boundaries than single-modality inputs. Rahman et al. [25] demonstrated that combining DWI, ADC, and CLAHE-enhanced DWI (eDWI) through a shared encoder improves segmentation performance, as each imaging sequence contributes complementary information. DWI highlights acute ischemic lesions, ADC quantifies diffusion restriction, and eDWI enhances lesion boundary visibility. Their framework further integrates the FSCA module with DenseNet-based encoders and Self-Organized Operational Neural Networks, achieving competitive segmentation performance. Nevertheless, the proposed architecture remains computationally demanding because of its relatively large network complexity, which may limit its practical deployment in time-sensitive clinical environments. Moreover, although the study demonstrates the effectiveness of multi-sequence MRI, integrating complementary imaging modalities with a computationally efficient UNet++-based architecture remains insufficiently explored.

2.5. Research Gaps and Contribution

Although considerable progress has been achieved in ischemic stroke lesion segmentation, several important limitations remain unresolved. These limitations motivate the development of the proposed framework and can be summarized into three main research gaps.

First, although UNet++ effectively reduces the semantic gap between encoder and decoder features, it still relies on conventional convolution blocks that are less effective in capturing multi-scale variations in ischemic stroke lesions [26]. Second, while the FSCA module has demonstrated promising performance in other medical imaging applications [14], its integration into UNet++ for ischemic stroke lesion segmentation has not been extensively investigated. Third, to the best of our knowledge, no previous study has evaluated the combined effect of Modified InceptionNeXt Blocks and FSCA within a unified UNet++ architecture for ischemic stroke lesion segmentation using multi-sequence MRI inputs. To address these research gaps, this study makes the following contributions:

- Develops an FSCA-InceptionNeXt-UNet++ architecture by integrating Modified InceptionNeXt Blocks for multi-scale feature extraction and the FSCA module for adaptive feature selection within the UNet++ framework.
- Evaluates three multi-sequence MRI input configurations (DWI, DWI+ADC, and DWI+ADC+eDWI) to investigate the contribution of complementary imaging modalities to segmentation performance.
- Conducts a comprehensive ablation study to analyze the individual and combined contributions of the proposed architectural components.

Collectively, previous studies indicate that attention mechanisms, multi-scale feature extraction, and multi-sequence MRI provide complementary advantages for ischemic stroke lesion segmentation. Building upon these findings, this study integrates these complementary strategies within a unified UNet++ framework to address the identified research gaps while maintaining computational efficiency.

3. Proposed Method

The overall workflow of the proposed ischemic stroke lesion segmentation framework is illustrated in Figure 1. The proposed framework consists of four main stages: data preparation, multi-sequence MRI construction, network architecture design, and model training and evaluation. Initially, the ISLES 2022 dataset undergoes image preprocessing, including resizing and CLAHE, to enhance lesion visibility and produce enhanced Diffusion-Weighted Imaging (eDWI). The original DWI, Apparent Diffusion Coefficient (ADC), and eDWI are subsequently concatenated to construct a three-channel multi-sequence input.

The resulting input is processed by the proposed UNet++ architecture, in which the conventional convolution blocks are replaced by Modified InceptionNeXt Blocks to improve multi-scale feature extraction while maintaining computational efficiency. To further enhance feature propagation between the encoder and decoder, the proposed architecture integrates the FSCA module into each skip connection, enabling adaptive feature selection during feature fusion. The network configuration is optimized through hyperparameter tuning, and the final model is evaluated using Dice Score, Intersection over Union (IoU), Precision, and Recall.

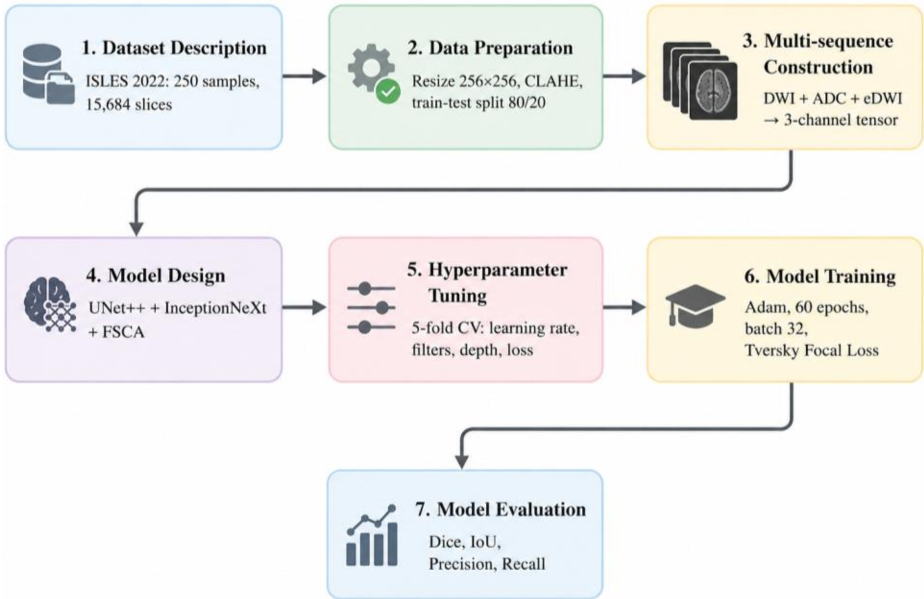

Figure 1. High-level pipeline of the proposed methodology for ischemic stroke lesion segmentation.

3.1. Dataset Description

This study employs the publicly available ISLES 2022 medical image dataset, which has been widely adopted for the development and evaluation of ischemic stroke lesion segmentation models. The dataset contains 400 brain MRI cases acquired from ischemic stroke patients, consisting of 250 publicly available cases for training and 150 private cases reserved for challenge evaluation. Because the private data are not publicly accessible, this study utilizes only the 250 public cases, following the experimental protocol adopted by Li et al. [27].

The ISLES 2022 dataset was selected because its characteristics closely reflect the challenges addressed by the proposed method. The dataset contains ischemic stroke lesions with volumes ranging from approximately 0.1 mL to over 100 mL, providing a rigorous benchmark for evaluating small-lesion segmentation. In addition, lesion morphology varies substantially across patients and stroke stages, while lesion boundaries often exhibit low contrast due to peri-infarct tissue. These characteristics provide an appropriate setting for evaluating the effectiveness of the proposed Modified InceptionNeXt Blocks in learning multi-scale lesion representations. Furthermore, lesion voxels typically constitute less than one percent of the total brain voxels, creating severe class imbalance that challenges both the proposed FSCA module and the Tversky Focal Loss during optimization. Finally, the availability of multiple MRI sequences, including DWI and ADC, enables the proposed framework to investigate whether complementary imaging information improves segmentation performance [16]. Collectively, these characteristics make ISLES 2022 an appropriate benchmark for evaluating multi-scale feature extraction, attention mechanisms, and imbalance-aware optimization strategies.

3.2. Data Preparation

Prior to model training, all MRI slices were resized to 256 × 256 pixels to provide a consistent input resolution while maintaining the original spatial characteristics of the medical images. Image resizing was performed using interpolation to ensure compatibility with the proposed network architecture without introducing substantial geometric distortion.

To improve lesion visibility, the DWI sequence was enhanced using CLAHE. CLAHE improves local image contrast by dividing each image into small regions (tiles) and performing adaptive histogram equalization with a predefined contrast limit, thereby enhancing lesion boundaries and local texture while suppressing excessive noise amplification [17]. The enhanced DWI (eDWI) was subsequently combined with the original DWI and ADC images to construct the multi-sequence input used by the proposed model. Following preprocessing, the dataset was divided into training and testing subsets using a patient-based splitting strategy to prevent data leakage. This strategy ensures that all slices belonging to the same patient are assigned exclusively to either the training or the testing subset, preventing the model from learning patient-specific characteristics that could artificially inflate the evaluation results.

Table 1. Characteristics of the ISLES 2022 dataset used in this study.

Parameter	Value
Total cases	250
Total slices	15,684
Training cases	200 (12,592 slices)
Testing cases	50 (3,092 slices)

As summarized in Table 1, the 250 publicly available ISLES 2022 cases were divided into 200 training cases and 50 testing cases, following the protocol reported by Li et al. [27]. This split produced a total of 15,684 MRI slices, comprising 12,592 training slices and 3,092 testing slices. An 80:20 patient-level split was adopted to provide sufficient training data while retaining an independent test set for unbiased performance evaluation. Hyperparameter optimization was performed exclusively on the 200 training cases using 5-fold cross-validation, during which the number of filters, network depth, and loss function were systematically evaluated. This strategy follows the validation protocol commonly adopted by nnU-Net for medical image segmentation, providing a reliable estimate of model generalization while reducing the risk of overfitting to a single validation split [8].

Within each fold, 160 cases (80% of the training set) were used for model training, while the remaining 40 cases (20%) were reserved for validation. After determining the optimal hyperparameter configuration, the final model was retrained using all 200 training cases and subsequently evaluated only once on the independent 50-case testing set. Consequently, the testing set was never involved in either hyperparameter optimization or cross-validation, ensuring an unbiased evaluation of the proposed method [8].

3.3. Model Design

The proposed model is built upon the UNet++ architecture, which was selected because its nested skip connections effectively reduce the semantic gap between encoder and decoder feature representations, leading to improved feature fusion and segmentation accuracy in medical imaging tasks [7]. Despite these advantages, the original UNet++ still relies on conventional convolution blocks, which may limit its ability to capture multi-scale lesion characteristics while efficiently suppressing irrelevant background information.

To address these limitations, the proposed architecture introduces two complementary modifications. First, the conventional convolution blocks are replaced with Modified InceptionNeXt Blocks to improve multi-scale feature extraction while maintaining computational efficiency. Second, the FSCA module is incorporated into each skip connection to selectively emphasize lesion-relevant features before encoder and decoder feature fusion. Together, these modifications are designed to enhance feature representation, improve information propagation across the network, and increase segmentation accuracy for ischemic stroke lesions with varying sizes, irregular morphologies, and low contrast.

3.3.1. Overall Architecture

The overall architecture of the proposed model is illustrated in Figure 2. The network adopts a hierarchical encoder-decoder structure based on UNet++, where the input tensor of 256 × 256 × 3 is processed through four encoder stages and four corresponding decoder stages connected by nested skip pathways. This design enables feature maps extracted at different semantic levels to be progressively fused, reducing the semantic discrepancy between encoder and decoder representations while preserving fine spatial information [7].

Unlike the original UNet++, each encoder and decoder stage employs a Modified InceptionNeXt Block instead of conventional convolution blocks. This modification enables the network to capture lesion features at multiple receptive fields while maintaining a relatively lightweight architecture. Between successive encoder stages, feature maps are downsampled using MaxPooling2D, whereas the decoder reconstructs spatial resolution through PixelShuffle upsampling, which effectively restores fine-grained structural details while minimizing checkerboard artifacts commonly associated with conventional transposed convolution.

To further improve feature fusion, the proposed architecture integrates an FSCA module into every skip connection. Rather than directly concatenating encoder and decoder features, the encoder features are first refined through attention before feature fusion. This design allows the network to suppress irrelevant background responses while preserving informative lesion representations, thereby improving the quality of information propagated to the decoder.

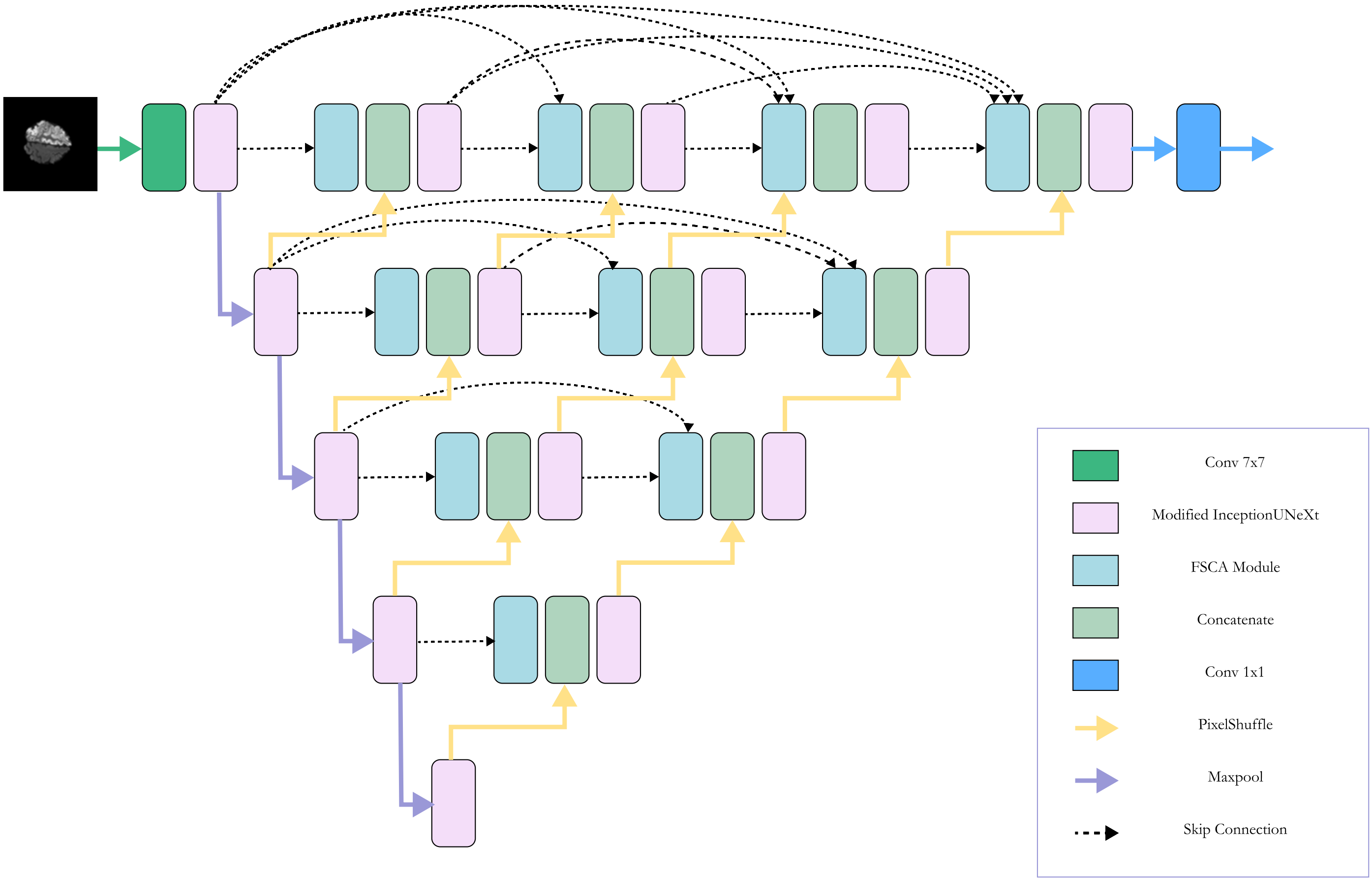

Figure 2. Proposed UNet++ architecture integrating Modified InceptionNeXt Blocks and the FSCA module.

As illustrated in Figure 2, the network follows the nested skip connection strategy of UNet++, where each encoder stage is connected to multiple decoder stages through progressively refined pathways. The feature extraction process begins from the input layer, proceeds through the encoder to learn increasingly abstract feature representations, and is subsequently reconstructed by the decoder using attention-guided skip connections. Finally, the output feature map at node X_4^0 is projected through a 1×1 convolution, followed by a sigmoid activation function, to generate the final binary segmentation probability map.

3.3.2. Modified InceptionNeXt Block

The Modified InceptionNeXt Block serves as the fundamental feature extraction module throughout both the encoder and decoder, as illustrated in Figure 3. Rather than employing the conventional convolution blocks used in the original UNet++, the proposed architecture replaces them with Modified InceptionNeXt Blocks to improve multi-scale feature extraction while maintaining computational efficiency. This design follows the Inception depthwise convolution principle introduced by Yu et al. [24] and adopts the parameter-efficient modifications proposed by Tagnamas et al. [14], with additional adjustments to accommodate the varying channel widths across the nested UNet++ architecture.

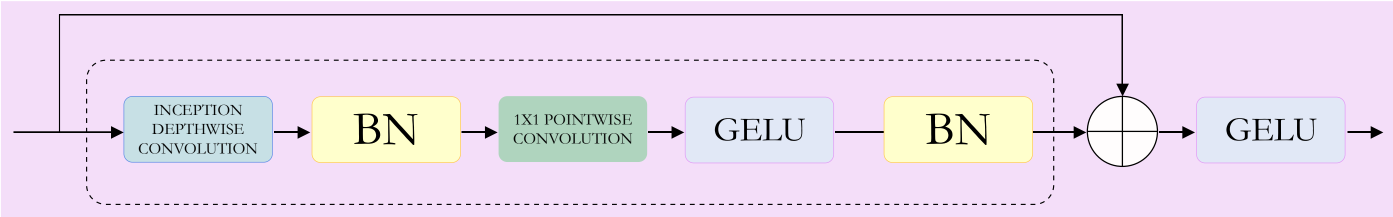

Figure 3. Structure of the proposed Modified InceptionNeXt Block.

Given an input feature map $X \in \mathbb{R}^{H\times W\times C_{in}}$, the input feature map is first normalized using batch normalization and then evenly divided into four groups along the channel dimension:

$$(X_1, X_2, X_3, X_4) = \mathrm{Split}(X), \qquad X_i \in \mathbb{R}^{H\times W\times \frac{C_{in}}{4}},\ \ i = 1, \ldots, 4 \tag{1}$$

where H and W denote the spatial dimensions of the feature map, C_{in} represents the number of input channels, and $\mathrm{Split}(\cdot)$ partitions the input tensor into four equally sized channel groups.

The first three channel groups are processed independently using grouped convolutions with different receptive fields, namely 3 × 3, 1 × 11, and 11 × 1 kernels, respectively, whereas the fourth group is preserved as an identity branch. This multi-branch design enables the network to simultaneously capture local texture information and elongated structural patterns while avoiding a substantial increase in computational cost. The resulting feature maps are concatenated along the channel dimension to produce the Inception depthwise representation:

$$X_{idw} = \mathrm{Concat}\left(\mathrm{GConv}^{g}_{3\times 3}(X_1), \mathrm{GConv}^{g}_{1\times 11}(X_2), \mathrm{GConv}^{g}_{11\times 1}(X_3), (X_4)\right) \tag{2}$$

where $\mathrm{Concat}(\cdot)$ denotes channel-wise concatenation and $\mathrm{GConv}^{g}_{k\times k}$ represents grouped convolution with kernel size $k \times k$ and g groups, defined as:

$$\mathrm{GConv}^{g}_{k\times k}(X) = \mathrm{Conv2D}_{k\times k, groups=g}(X) \tag{3}$$

Using grouped convolutions reduces the number of trainable parameters while preserving sufficient feature diversity within each channel group, making the block well suited for medical image segmentation tasks with relatively limited training data.

The concatenated feature map is subsequently normalized, projected using a 1 × 1 pointwise convolution, activated by the Gaussian Error Linear Unit (GELU), and combined with a residual shortcut to facilitate gradient propagation during training:

$$Y = \mathrm{GELU}\left(\mathrm{Conv}_{1\times 1}\left(\mathrm{BN}(X_{idw})\right)\right) + P(X) \tag{4}$$

where Y denotes the output feature map, $\mathrm{BN}(\cdot)$ is batch normalization, and $P(X)$ represents the shortcut connection defined as:

$$P(X) = \begin{cases} X, & C_{in} = C_{out}, \\ \mathrm{Conv}_{1\times 1}(X), & C_{in} \neq C_{out}. \end{cases} \tag{5}$$

When the input and output channel dimensions are identical, the shortcut is implemented as an identity mapping. Otherwise, a 1 × 1 projection convolution is applied to align the channel dimensions before residual addition.

3.3.3. Spatial-aware Channel Attention (SCA)

To improve the quality of feature propagation between the encoder and decoder, the proposed architecture employs a SCA module, following the attention mechanism introduced by Tagnamas et al. [14]. Rather than directly propagating encoder features through skip connections, the SCA module selectively emphasizes lesion-relevant responses while suppressing redundant background activations before feature fusion. This attention mechanism is particularly beneficial for ischemic stroke lesion segmentation, where lesions are often small, exhibit irregular morphology, and are surrounded by extensive normal tissue.

As illustrated in Figure 4, the SCA module consists of three complementary attention paths: a spatial-attention branch and two cross-dimensional branches that model the interactions between the channel dimension and the height and width dimensions, respectively. By combining these complementary attention mechanisms, the module enhances both local lesion representation and contextual spatial information before feature fusion.

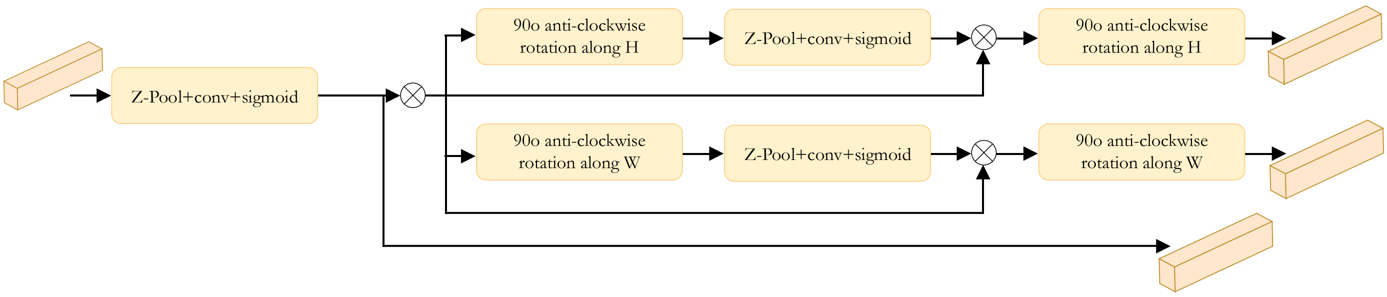

Figure 4. Structure of the Spatial-aware Channel Attention (SCA) module.

For the spatial-attention branch, the input feature map $X \in \mathbb{R}^{H\times W\times C}$ is first processed using channel-wise average pooling and max pooling. The pooled features are concatenated, followed by a 7 × 7 convolution, batch normalization, and a sigmoid activation to generate the spatial attention map:

$$X_s = \sigma\left(\mathrm{BN}\left(\mathrm{Conv}_{7\times 7}\left(\mathrm{ZPool}_C(X)\right)\right)\right) \odot X \tag{6}$$

where X_s enotes the spatially refined feature map, $\mathrm{ZPool}_C(\cdot)$ represents the concatenation of channel-wise average pooling and max pooling, $\sigma(\cdot)$ is the sigmoid activation function, and $\odot$ denotes element-wise multiplication.

The 7 × 7 convolution kernel is retained in the proposed implementation because preliminary hyperparameter tuning demonstrated that its larger receptive field provides better validation performance for detecting the relatively small lesion regions present in the ISLES 2022 dataset. The spatially refined feature map is subsequently processed by two cross-dimensional attention branches to capture complementary dependencies along the height and width dimensions.

For the height branch, the feature map is first permuted so that the height dimension becomes the pooling axis, followed by Z-pooling, convolution, and self-gated attention before being restored to its original orientation:

$$A_h = \mathrm{BN}\left(\mathrm{Conv}_{7\times 7}\left(\mathrm{ZPool}_H\left(R_h(X_s)\right)\right)\right),\ \ X_h = R_h^{-1}\left(A_h \odot \sigma(A_h)\right), \tag{7}$$

where $R_h(\cdot)$ denotes the permutation operation that aligns the channel dimension with the height dimension, and $R_h^{-1}(\cdot)$ restores the original feature arrangement. Similarly, the width branch performs the same operations after aligning the channel dimension with the width axis:

$$A_w = \mathrm{BN}\left(\mathrm{Conv}_{7\times 7}\left(\mathrm{ZPool}_W\left(R_w(X_s)\right)\right)\right),\ \ X_w = R_w^{-1}\left(A_w \odot \sigma(A_w)\right), \tag{8}$$

where $R_w(\cdot)$ and $R_w^{-1}(\cdot)$ denote the forward and inverse permutation operations for the width dimension, respectively. Finally, the outputs of the three attention branches are combined by simple averaging to produce the refined feature representation:

$$X_{SCA} = \frac{1}{3}(X_s + X_h + X_w) \tag{9}$$

The resulting feature map X_{SCA} preserves complementary spatial and cross-dimensional information, providing more discriminative encoder features before they are fused with decoder features through the proposed FSCA module.

3.3.4. Feature-Scale Channel Attention (FSCA)

Following the SCA module, the proposed architecture employs FSCA to improve feature fusion between the encoder and decoder. Unlike conventional UNet++ skip connections that directly concatenate encoder and decoder features, the proposed framework first refines encoder features through attention before feature fusion. This strategy allows lesion-relevant information to be preserved while suppressing irrelevant background responses, resulting in more discriminative feature representations for subsequent decoding [14].

As illustrated in Figure 5, the FSCA module receives encoder features E_i and decoder features D_i from the corresponding resolution level. The two feature maps are first concatenated along the channel dimension and subsequently processed by the SCA module to generate an attention-refined representation. The refined features are then passed through a 3 × 3 convolution, batch normalization, and GELU activation before being processed by the Modified InceptionNeXt Block. This sequential refinement enables the network to selectively fuse complementary encoder and decoder information while preserving multi-scale contextual features.

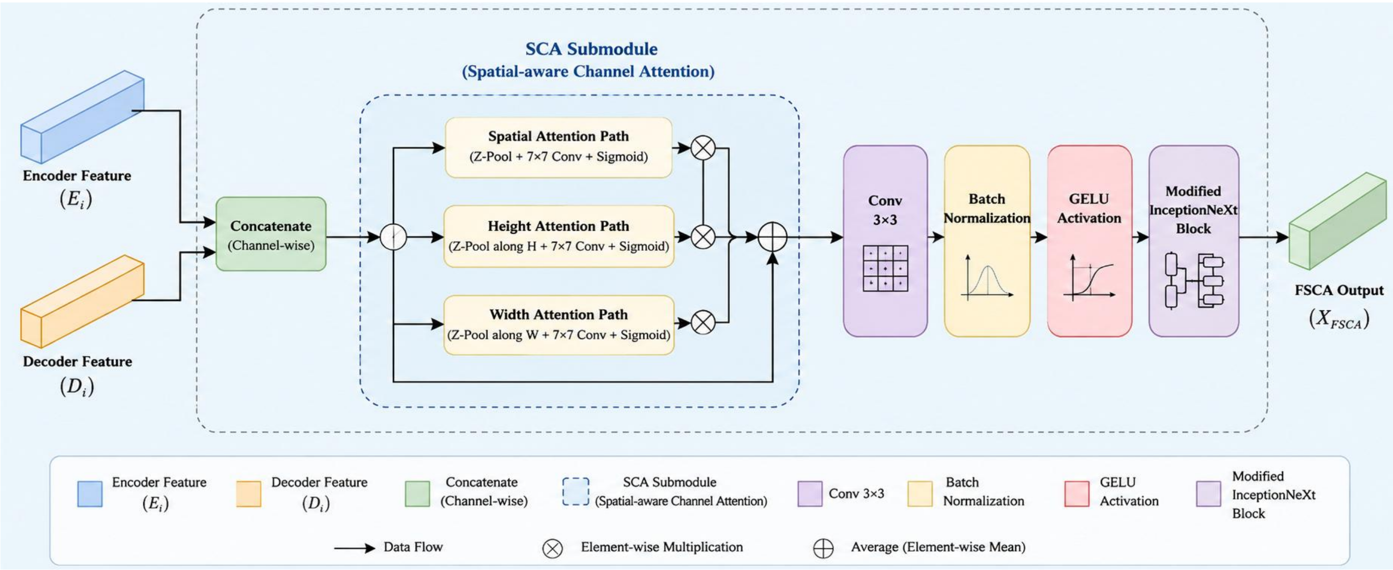

Figure 5. Structure of the proposed Feature-Scale Channel Attention (FSCA) module.

The overall operation of the FSCA module can be expressed as:

$$X_{FSCA} = \mathrm{IncBlock}\left(\mathrm{GELU}\left(\mathrm{BN}\left(\mathrm{Conv}_{3\times3}\left(\mathrm{SCA}\left(\mathrm{Concat}(E_i, D_i)\right)\right)\right)\right)\right) \tag{10}$$

where E_i and D_i denote the encoder and decoder feature maps at the i-th resolution level, respectively; $\mathrm{Concat}(\cdot)$ represents channel-wise concatenation; $\mathrm{SCA}(\cdot)$ is the Spatial-aware Channel Attention module defined in Equations (6)–(9); and $\mathrm{IncBlock}(\cdot)$ denotes the Modified InceptionNeXt Block described in Equations (1)–(5).

By integrating attention-guided feature refinement with multi-scale feature extraction, the FSCA module enables the decoder to reconstruct segmentation maps from more informative feature representations, thereby improving lesion delineation, particularly for small and low-contrast ischemic stroke lesions.

3.3.5. Decoder and Output Layer

After feature refinement through the FSCA module, the decoder progressively reconstructs the segmentation map by integrating high-level semantic information with fine-grained spatial details obtained from the encoder. The hierarchical decoder follows the nested

UNet++ topology, enabling feature maps from multiple semantic levels to be fused progressively throughout the reconstruction process.

At each decoder stage, the feature map from the deeper level is first upsampled using PixelShuffle, which increases the spatial resolution while preserving structural continuity and reducing checkerboard artifacts commonly associated with transposed convolution. The upsampled features are subsequently fused with the corresponding encoder features that have been refined by the FSCA module. The fused representation is then processed using the Modified InceptionNeXt Block, allowing multi-scale contextual information to be extracted before being propagated to the next decoder stage.

The final decoder output is generated at node X_0^4, where a 1×1 pointwise convolution projects the multi-channel feature representation into a single-channel probability map. A sigmoid activation function is subsequently applied to estimate the probability of each pixel belonging to the ischemic stroke lesion, thereby producing the final binary segmentation map. The principal architectural configuration of the proposed model is summarized in Table 2.

Table 2. Model configuration of the proposed architecture.

Component	Configuration
Normalization	Batch Normalization
Activation Function	GELU
Output Activation	Sigmoid
Upsampling	PixelShuffle
Pooling	MaxPooling2D

The proposed decoder is designed to balance segmentation accuracy and computational efficiency. The combination of attention-guided feature fusion, multi-scale feature extraction, and lightweight decoder operations enables the network to preserve detailed lesion boundaries while maintaining a relatively compact architecture suitable for practical medical image segmentation applications.

3.4. Hyperparameter Tuning

Hyperparameter tuning was conducted to identify the optimal configuration of the proposed model before final training. Rather than relying on default settings, several architectural and optimization parameters were systematically explored because they directly influence feature representation, convergence behavior, and segmentation performance. The investigated hyperparameters included the initial number of filters, network depth, learning rate, and loss function. The tuning process was performed exclusively on the training set using 5-fold cross-validation, following the data partitioning strategy described in Section 3.2. For each hyperparameter configuration, the model was trained independently on four folds and validated on the remaining fold. The average validation performance across all folds was then used to determine the optimal configuration. Throughout this process, the independent testing set was never accessed, ensuring that no information from the testing data influenced model selection. The hyperparameter search was designed to balance segmentation accuracy, model complexity, and generalization capability. The optimal configuration obtained from this procedure was subsequently used for the final model training, while the quantitative comparison of the evaluated hyperparameter settings is presented in Section 4.2–4.5.

3.5. Model Training

After determining the optimal hyperparameter configuration, the proposed model was retrained using the complete training set comprising 200 patient cases. Three input configurations were investigated throughout the experiments, namely DWI, DWI + ADC, and DWI + ADC + eDWI, to evaluate the contribution of complementary MRI sequences to lesion segmentation. Each configuration was processed independently using the same network architecture and training protocol to ensure a fair comparison.

The proposed network was implemented using TensorFlow 2.19.0 and optimized with the Adam optimizer [28] using a learning rate of 0.001. Model training was performed for 60 epochs with a batch size of 32, while the model weights achieving the lowest validation loss were retained through model checkpointing. No explicit data augmentation was applied beyond the CLAHE-based enhancement described in Section 3.2, allowing the effectiveness of the proposed architectural modifications to be evaluated without additional augmentation effects.

To address the severe class imbalance commonly encountered in ischemic stroke lesion segmentation, the network was optimized using Tversky Focal Loss [18]. The loss function was configured with $\alpha=\beta=0.5$, $\gamma=2$, and a smoothing constant of 1×10^{-7}, providing balanced optimization between false-positive and false-negative predictions. This configuration is well suited for medical image segmentation tasks in which lesion regions occupy only a small proportion of the entire image.

3.6. Model Evaluation

The final model was evaluated using the independent testing set described in Section 3.2. Performance was assessed using four widely adopted segmentation metrics: Dice Score, Intersection over Union (IoU), Precision, and Recall. These metrics provide complementary perspectives on segmentation quality. Dice Score and IoU measure the overlap between the predicted segmentation and the ground truth, whereas Precision and Recall quantify the trade-off between false-positive and false-negative predictions. Considering all four metrics together provides a comprehensive assessment of segmentation performance, particularly for highly imbalanced medical image datasets.

To further evaluate the effectiveness of the proposed architecture, the final model was compared with several representative deep learning approaches previously reported for ischemic stroke lesion segmentation, including DenseNet121-UNet, DenseNet121-UNet++, DeepLabV3, and SelfONN-CSCA-UNet [16]. In addition, an ablation study was conducted to investigate the individual and combined contributions of the Modified InceptionNeXt Block and the FSCA module to the overall segmentation performance.

4. Results and Discussion

This section presents the experimental results of the proposed UNet++ architecture integrated with the Modified InceptionNeXt Block and FSCA for ischemic stroke lesion segmentation. The experiments were conducted to evaluate the effectiveness of the proposed architecture through a series of hyperparameter tuning experiments, followed by the evaluation of the final model using the optimal configuration. Model performance was assessed using the Dice Score, Intersection over Union (IoU), Precision, and Recall metrics, and compared with several representative deep learning methods for ischemic stroke lesion segmentation. In addition, an ablation study was performed to analyze the individual and combined contributions of the proposed architectural components.

4.1. Experimental Setup

All experiments were implemented using the Python programming language with TensorFlow 2.19.0 as the deep learning framework. Model training and evaluation were performed on the Google Colaboratory platform using an NVIDIA A100 GPU equipped with 40 GB of VRAM and 6,912 CUDA cores, providing sufficient computational resources for both hyperparameter optimization and model evaluation. Following the methodology described in Section 3, the experiments were organized into four stages. The first stage investigated the effects of the learning rate, initial number of filters, network depth, and loss function to determine the optimal model configuration. The second stage evaluated different multi-sequence MRI input configurations. The third stage compared the proposed model with representative segmentation methods reported in the literature. Finally, an ablation study was conducted to evaluate the individual and combined contributions of the Modified InceptionNeXt Block and the FSCA module.

4.2. Learning Rate Experiment

The learning rate is one of the most influential optimization hyperparameters, as it determines the step size used to update the network parameters during training. An excessively large learning rate may prevent convergence, whereas a very small learning rate often leads to slow convergence and suboptimal optimization. To identify the most appropriate configuration for the proposed model, three learning rates (0.01, 0.001, and 0.0001) were evaluated while maintaining all other training settings unchanged.

Table 3 summarizes the experimental results. Among the evaluated configurations, a learning rate of 0.001 achieved the best overall performance, yielding a Dice Score of 0.8594 and an IoU of 0.8307, while maintaining a good balance between Precision (0.9412) and Recall (0.8808). In contrast, a learning rate of 0.01 resulted in substantially lower segmentation performance, indicating that the optimization process became unstable due to excessively large parameter updates. Conversely, although a learning rate of 0.0001 produced the highest Recall (0.8906), both the Dice Score and IoU were lower than those obtained with 0.001, suggesting that the slower optimization process limited convergence within the allocated training epochs.

Therefore, a learning rate of 0.001 was selected for all subsequent experiments because it provided the best trade-off between convergence stability and segmentation accuracy. This observation is also consistent with previous studies that commonly adopt 0.001 as an effective default learning rate for Adam-based optimization in deep learning applications [28].

Table 3. Experimental results for different learning rates.

Learning Rate	Dice Score ↑	IoU ↑	Precision ↑	Recall ↑
0.01	0.7899	0.7658	0.9619	0.7911
0.001	**0.8594**	**0.8307**	0.9412	0.8808
0.0001	0.8349	0.8091	**0.9440**	**0.8906**

4.3. Initial Feature Count Experiment

The number of initial convolution filters determines the feature extraction capacity of the network and directly affects both representational power and computational complexity. A larger number of filters generally enables richer feature representations but also increases the number of trainable parameters and the risk of overfitting. To investigate this trade-off, three initial filter configurations (8, 16, and 32) were evaluated while keeping all other training settings unchanged.

Table 4 summarizes the experimental results. Increasing the number of filters consistently improved segmentation performance across all evaluation metrics. The configuration with 32 initial filters achieved the highest Dice Score (0.8742) and IoU (0.8464) while also providing the highest Precision (0.9440) and Recall (0.8906). In comparison, the 8-filter configuration produced lower performance, suggesting that its limited representational capacity was insufficient to capture the subtle appearance of ischemic stroke lesions.

The gradual improvement observed as the number of filters increased indicates that richer feature representations are beneficial for distinguishing small lesion regions from surrounding healthy tissue. Nevertheless, further increasing the number of filters would substantially increase computational cost and model complexity without guaranteeing proportional performance gains, particularly given the relatively limited size of the ISLES 2022 dataset [8]. Therefore, 32 initial filters were selected as the optimal configuration for the proposed model.

Table 4. Experimental results for different initial feature count

Initial Filters	Dice Score ↑	IoU ↑	Precision ↑	Recall ↑
8	0.8594	0.8307	0.9321	0.8847
16	0.8634	0.8353	0.9412	0.8808
32	**0.8742**	**0.8464**	**0.9440**	**0.8906**

4.4. Architectural Depth Experiment

Network depth determines the hierarchical representation capability of the proposed architecture. Increasing the number of encoder-decoder levels generally enables the network to capture more abstract semantic features, although excessively deep architectures may increase computational complexity and the risk of overfitting. To determine an appropriate balance, the proposed model was evaluated using network depths of 3, 4, and 5.

The experimental results are presented in Table 5. Overall, deeper architectures produced better segmentation performance. The 5-level configuration achieved the highest Dice Score (0.8715), IoU (0.8431), and Precision (0.9456), indicating that the additional encoder-decoder level improved the network's ability to learn discriminative lesion representations.

Although the improvement over the 4-level architecture was relatively modest, it remained consistent across the primary segmentation metrics.

Table 5. Experimental results for different architectural depths.

Network Depth	Dice Score ↑	IoU ↑	Precision ↑	Recall ↑
3	0.8634	0.8353	0.9412	0.8808
4	0.8649	0.8350	0.9256	0.8842
5	**0.8715**	**0.8431**	**0.9456**	**0.8852**

These findings suggest that increasing network depth enhances the extraction of high-level semantic features required for ischemic stroke lesion segmentation. However, substantially deeper architectures would introduce additional parameters and computational cost while offering diminishing performance improvements [7], [8]. Therefore, the 5-level configuration was adopted in the subsequent experiments.

4.5. Loss Function Experiment

Selecting an appropriate loss function is particularly important for ischemic stroke lesion segmentation because lesion pixels occupy only a small proportion of the entire image. Conventional loss functions often struggle to balance false-positive and false-negative predictions under severe class imbalance. To investigate the influence of the optimization objective, three loss functions were evaluated: Dice Loss, Binary Cross-Entropy (BCE), and Tversky Focal Loss.

As shown in Table 6, Tversky Focal Loss achieved the best overall segmentation performance, producing the highest Dice Score (0.8715) and IoU (0.8431) while maintaining a balanced Precision (0.9456) and Recall (0.8852). In contrast, Dice Loss achieved perfect Precision but substantially lower Recall, indicating that the model tended to predict only the most confident lesion pixels while missing a considerable portion of the lesion area. Similarly, the BCE objective produced the lowest overall performance, reflecting its limited ability to handle the severe foreground-background imbalance present in the ISLES 2022 dataset.

The superior performance of Tversky Focal Loss can be attributed to its ability to simultaneously address class imbalance and emphasize difficult-to-segment lesion regions through focal weighting [18]. This property enables a better balance between Precision and Recall, making it particularly suitable for ischemic stroke lesion segmentation where lesion boundaries are often small, irregular, and exhibit low contrast. Accordingly, Tversky Focal Loss was adopted for all subsequent experiments.

Table 6. Experimental results for different loss functions.

Loss Function	Dice Score ↑	IoU ↑	Precision ↑	Recall ↑
Dice Loss	0.7039	0.7039	**1.0000**	0.7039
BCE	0.6103	0.5621	0.9240	0.5293
Tversky Focal Loss	**0.8715**	**0.8431**	0.9456	**0.8852**

4.6. Model Ablation Study

An ablation study was conducted to investigate the individual and combined contributions of the proposed architectural components to ischemic stroke lesion segmentation. Four model configurations were evaluated: the baseline UNet++, UNet++ + FSCA, UNet++ + Modified InceptionNeXt Block, and the complete model integrating both FSCA and the Modified InceptionNeXt Block. All configurations were evaluated under the same experimental settings using the independent testing set, with segmentation performance assessed using the Dice Score, Intersection over Union (IoU), Precision, and Recall metrics. The experimental results are summarized in Table 7.

The baseline UNet++ achieved a Dice Score of 0.8614, demonstrating that the nested skip connections effectively preserve spatial information for ischemic stroke lesion segmentation [7]. This baseline provides an appropriate reference for evaluating the contribution of each proposed architectural modification. Introducing the FSCA module alone slightly reduced performance across all evaluation metrics, yielding a Dice Score of 0.8426. This observation suggests that attention mechanisms alone are insufficient to improve segmentation performance when the underlying feature representations remain unchanged. Since the original UNet++ encoder relies on conventional convolution blocks, the extracted features may not contain sufficiently rich multi-scale information for the attention mechanism to exploit effectively. Consequently, the FSCA module has limited capacity to enhance feature discrimination when applied in isolation.

Table 7. Ablation study results of the proposed architecture.

Model Configuration	Dice Score ↑	IoU ↑	Precision ↑	Recall ↑
UNet++	0.8614	0.8326	0.9215	0.8974
UNet++ + FSCA	0.8426	0.8136	0.9069	0.8912
UNet++ + Modified InceptionNeXt Block	0.7235	0.6919	0.7630	0.9068
UNet++ + FSCA + Modified InceptionNeXt Block	**0.8742**	**0.8464**	**0.9440**	**0.8906**

Replacing the conventional convolution blocks with the Modified InceptionNeXt Block alone resulted in a more pronounced performance degradation, reducing the Dice Score to 0.7235. Although the Recall increased slightly (0.9068), Precision decreased substantially (0.7630), indicating that the model became more sensitive to lesion regions while simultaneously generating a considerably larger number of false-positive predictions. This behavior suggests that the multi-scale branches introduced by the Modified InceptionNeXt Block successfully capture richer contextual information but also amplify irrelevant background responses. Without an effective feature selection mechanism, these noisy feature representations propagate through the decoder and negatively affect segmentation quality, particularly for small and low-contrast lesions [14].

The best performance was achieved when both the FSCA module and the Modified InceptionNeXt Block were integrated into the proposed architecture, producing a Dice Score of 0.8742, IoU of 0.8464, Precision of 0.9440, and Recall of 0.8906. Compared with the baseline UNet++, the complete model consistently improved the overlap-based metrics while maintaining a balanced trade-off between Precision and Recall. More importantly, compared with the configuration employing only the Modified InceptionNeXt Block, Precision increased markedly from 0.7630 to 0.9440, indicating that the FSCA module effectively suppressed the false-positive responses introduced by the multi-scale feature extraction process.

Overall, the ablation results demonstrate that the proposed architectural components perform complementary rather than independent roles. The Modified InceptionNeXt Block enriches feature representations by capturing lesion characteristics at multiple receptive fields, whereas the FSCA module selectively emphasizes informative features while suppressing irrelevant background activations [14], [24]. Consequently, neither component alone is sufficient to maximize segmentation performance; however, their integration enables the network to simultaneously achieve rich multi-scale feature extraction and effective feature selection. These findings are consistent with the design hypothesis proposed in this study, confirming that combining multi-scale feature learning with attention-guided feature refinement provides a more effective strategy for ischemic stroke lesion segmentation than applying either mechanism individually.

5. Discussion

5.1. Overall Segmentation Performance

Based on the hyperparameter tuning experiments described in Section 4, the optimal configuration was selected for training the final segmentation model. The resulting configuration represents the best combination of architectural and optimization parameters identified during the experimental evaluation and was consistently adopted for all subsequent quantitative comparisons and qualitative analyses. The complete configuration is summarized in Table 8.

Using the optimal configuration, the proposed model was evaluated on the ISLES 2022 dataset using the combined DWI, ADC, and CLAHE-enhanced DWI (eDWI) input configuration. These complementary MRI sequences provide different types of diagnostic information: DWI highlights acute ischemic lesions, ADC characterizes diffusion restriction, and eDWI improves lesion visibility by enhancing local image contrast. Their integration enables the network to exploit richer anatomical and pathological information, thereby improving lesion representation during feature learning.

Table 8. Final hyperparameter configuration of the proposed model.

Hyperparameter	Value
Batch Size	32
Initial Filters	32
Network Depth	5
Training Epochs	60
Cross-Validation	5-fold
Optimizer	Adam
Learning Rate	0.001
Loss Function	Tversky Focal Loss

Table 9. Segmentation performance of the proposed model on the ISLES 2022 dataset.

Configuration/ Metric	Value
Input Modality	DWI + ADC + eDWI
Dice Score	0.8742
IoU	0.8464
Precision	0.9440
Recall	0.8906
Parameters	11,366,925

As summarized in Table 9, the proposed model achieved a Dice Score of 0.8742 and an IoU of 0.8464, indicating excellent agreement between the predicted segmentation and the expert annotations. The model also achieved a Precision of 0.9440 and a Recall of 0.8906, demonstrating its ability to accurately delineate lesion boundaries while successfully detecting the majority of ischemic stroke lesions. From a clinical perspective, maintaining high Recall is particularly important because missed lesion regions (false negatives) may delay diagnosis or lead to an underestimation of stroke severity, whereas a high Precision reduces unnecessary false-positive predictions that could increase the clinician's workload. The proposed model maintains a favorable balance between these two objectives, making it well suited for automated ischemic stroke lesion segmentation.

The proposed architecture contains 11.36 million trainable parameters, reflecting a relatively lightweight network compared with many recent deep learning-based segmentation models. This moderate model complexity is achieved through the parameter-efficient design of the Modified InceptionNeXt Block together with the selective feature refinement provided by the FSCA module, allowing the model to achieve competitive segmentation performance without substantially increasing computational cost.

In addition to the quantitative evaluation, a qualitative analysis was conducted to assess the visual quality of the predicted segmentation masks. Representative segmentation examples are presented in Figure 6, including the original DWI, ADC, and CLAHE-enhanced DWI (eDWI) images together with the corresponding expert annotation, predicted segmentation, and overlay visualization.

Based on the qualitative results shown in Figure 6, the proposed model accurately identifies the lesion location and produces segmentation boundaries that closely follow the expert annotations. The overlay visualization further demonstrates a high degree of spatial correspondence between the predicted lesion regions and the ground truth. Although minor discrepancies remain around several small or low-contrast lesion boundaries, the overall segmentation quality is consistent across the evaluated examples. These observations complement the quantitative results and indicate that integrating multi-sequence MRI information with the Modified InceptionNeXt Block and the FSCA module enables the proposed architecture to generate accurate and clinically meaningful ischemic stroke lesion segmentations.

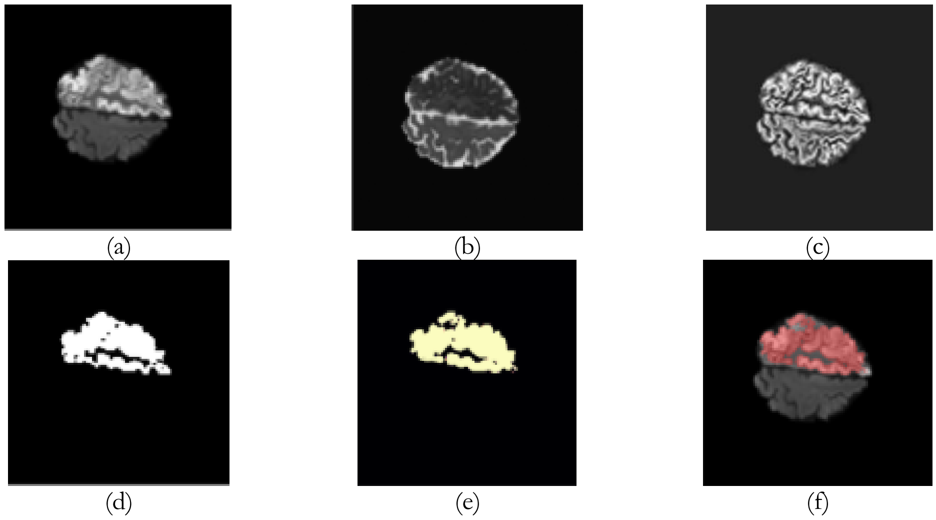

Figure 6. Representative qualitative segmentation results obtained using the proposed model on the ISLES 2022 dataset: (**a**) DWI image; (**b**) ADC image, (**c**) CLAHE-enhanced DWI (eDWI); (**d**) expert-annotated ground truth; (**e**) predicted segmentation; and (**f**) overlay visualization.

5.2. Comparison with Previous Methods

To further evaluate the effectiveness of the proposed architecture, the FSCA-InceptionNeXt-UNet++ model was compared with several representative deep learning approaches for ischemic stroke lesion segmentation, including DenseNet121-UNet, DenseNet121-UNet++, DeepLabV3, and SelfONN-CSCA-UNet [16]. The comparison considered both segmentation performance and computational efficiency using three input configurations: DWI, DWI + ADC, and DWI + ADC + eDWI. The experimental results are summarized in Table 10.

The proposed FSCA-InceptionNeXt-UNet++ demonstrates a favorable balance between segmentation performance and computational efficiency. Using the DWI + ADC + eDWI input configuration, it achieved a Dice Score of 0.8742 and an IoU of 0.8464, outperforming DenseNet121-UNet, DenseNet121-UNet++, and DeepLabV3, while achieving performance comparable to SelfONN-CSCA-UNet. In addition, the proposed model achieved the highest Recall (0.8906) among all compared methods while maintaining a high Precision (0.9440), indicating its ability to detect a larger proportion of ischemic stroke lesions without substantially increasing false-positive predictions.

From a clinical perspective, high Recall is particularly important because missed lesion regions (false negatives) may delay treatment decisions or underestimate stroke severity. Although the proposed model does not achieve the highest Precision, the difference is relatively small compared with the best-performing method (0.9440 versus 0.9501). This trade-off is clinically acceptable because false-positive regions can subsequently be reviewed by clinicians, whereas missed lesions may have considerably greater clinical consequences [1], [3].

The proposed architecture also demonstrates superior computational efficiency. As shown in Table 11, it contains only 11.36 million trainable parameters, the fewest among all compared models (13.61–39.63 million parameters), while achieving the fastest inference time (0.0164 s) for the three-modality input configuration. This efficiency is primarily attributed to the Modified InceptionNeXt Block, which performs parameter-efficient multi-scale feature extraction, and the FSCA module, which selectively refines encoder-decoder feature fusion without introducing substantial computational overhead [14], [24]. These results demonstrate that the proposed architecture effectively balances segmentation accuracy, computational efficiency, and clinical applicability. By combining efficient multi-scale feature extraction, attention-guided feature refinement, and complementary multi-sequence MRI information, the proposed model delivers accurate lesion segmentation while remaining suitable for practical deployment in resource-constrained clinical environments.

Table 10. Comparison of segmentation performance and computational efficiency with previous methods.

Model	Modality	Parameters (M) ↓	Inference Time (s) ↓	Dice Score ↑	IoU ↑	Precision ↑	Recall ↑
DenseNet121-UNet [16]	DWI	13.61	0.0250	0.6898	–	–	–
	DWI + ADC		0.0300	0.8354	–	–	–
	DWI + ADC + eDWI		0.0350	0.8643	0.8350	0.9501	0.8715
DenseNet121-UNet++ [16]	DWI	30.07	0.0280	0.7989	–	–	–
	DWI + ADC		0.0330	0.8407	–	–	–
	DWI + ADC + eDWI		0.0390	0.8635	0.8341	0.9443	0.8761
DeepLabV3 [16]	DWI	39.63	0.0300	0.7506	–	–	–
	DWI + ADC		0.0355	0.8142	–	–	–
	DWI + ADC + eDWI		0.0403	0.8440	0.8138	0.9360	0.8540
SelfONN-CSCA-UNet [16]	DWI	21.60	0.0276	0.8388	–	–	–
	DWI + ADC		0.0335	0.8586	–	–	–
	DWI + ADC + eDWI		0.0354	0.8749	0.8463	0.9487	0.8857
FSCA-Inception-NeXt-UNet++ (Proposed)	DWI	**11.36**	**0.0098**	**0.8577**	–	–	–
	DWI + ADC		**0.0099**	**0.8541**	–	–	–
	DWI + ADC + eDWI		**0.0164**	**0.8742**	**0.8464**	**0.9440**	**0.8906**

6. Conclusions

This study proposed an enhanced UNet++ architecture that integrates the Modified InceptionNeXt Block and FSCA to improve ischemic stroke lesion segmentation from multi-sequence MRI images. By combining complementary information from DWI, ADC, and CLAHE-enhanced DWI (eDWI) with multi-scale feature extraction and attention-guided feature refinement, the proposed framework effectively addresses several key challenges in ischemic stroke segmentation, including small lesion size, irregular lesion morphology, low image contrast, and severe class imbalance.

The experimental results demonstrate that the proposed architecture achieves competitive segmentation performance while maintaining high computational efficiency. The ablation study further confirms that the Modified InceptionNeXt Block and FSCA perform complementary rather than independent roles. The Modified InceptionNeXt Block enriches multi-scale feature representations, whereas the FSCA module selectively emphasizes informative lesion features and suppresses irrelevant background responses during encoder-decoder feature fusion. Their integration enables the network to achieve more accurate and robust lesion segmentation than either component alone.

Beyond segmentation accuracy, the proposed model also offers practical advantages through its lightweight architecture and fast inference, making it suitable for deployment in resource-constrained clinical environments where timely decision support is essential. These findings suggest that combining efficient multi-scale feature learning with attention-guided feature refinement provides an effective strategy for developing accurate and computationally efficient medical image segmentation models. Nevertheless, this study has several limitations. The experiments were conducted using only the publicly available ISLES 2022 dataset, and the proposed framework was implemented using a 2D segmentation strategy without external validation on independent datasets. Future work will therefore focus on extending the proposed architecture to 3D volumetric segmentation, validating its generalization capability on additional benchmark datasets such as ISLES 2015 and ATLAS, performing lesion-wise detection analysis, and further optimizing the model for real-time clinical deployment.

Author Contributions: Conceptualization: M.H.N. and W.; Methodology, M.H.N., W., and H.I.D.; Software, M.H.N.; Validation, M.H.N., W., and H.I.D.; Formal Analysis, M.H.N.; Investigation: M.H.N.; Resources: M.H.N.; Data Curation: M.H. N.; Writing – Original Draft Preparation, M.H.N.; Writing – Review and Editing: W. and H.I.D.; Visualization: M.H.N.; Supervision: W.; Project Administration: W. All authors have read and approved the final manuscript and agreed to be accountable for all aspects of the work.

Funding: This research received no external funding.

Data Availability Statement: The ISLES 2022 dataset used in this study is publicly available through the ISLES challenge repository at https://isles22.grand-challenge.org/. The dataset comprises 250 publicly available brain MRI cases from ischemic stroke patients, including Diffusion-Weighted Imaging (DWI) and Apparent Diffusion Coefficient (ADC) sequences, along with corresponding expert-annotated ground truth segmentation masks. No new data were generated in this study. The preprocessed data are not publicly shared due to the substantial preprocessing effort involved and the terms of use of the original dataset.

Acknowledgments: The authors used generative artificial intelligence solely to assist with English language editing and improve the clarity of the manuscript. All scientific content, methodology, data analysis, interpretation of results, and final manuscript revisions were performed and verified by the authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] W. J. Powers *et al.*, "Guidelines for the Early Management of Patients With Acute Ischemic Stroke: 2019 Update to the 2018 Guidelines for the Early Management of Acute Ischemic Stroke: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke ," *Stroke*, vol. 50, no. 12, Dec. 2019, doi: 10.1161/STR.0000000000000211.

[2] S. S. Virani *et al.*, "Heart Disease and Stroke Statistics—2020 Update: A Report From the American Heart Association," *Circulation*, vol. 141, no. 9, Mar. 2020, doi: 10.1161/CIR.0000000000000757.

[3] J. B. Fiebach *et al.*, "CT and Diffusion-Weighted MR Imaging in Randomized Order," *Stroke*, vol. 33, no. 9, pp. 2206–2210, Sep. 2002, doi: 10.1161/01.STR.0000026864.20339.CB.

[4] G. W. Albers *et al.*, "Thrombectomy for Stroke at 6 to 16 Hours with Selection by Perfusion Imaging," *N. Engl. J. Med.*, vol. 378, no. 8, pp. 708–718, Feb. 2018, doi: 10.1056/NEJMoa1713973.

[5] O. Maier *et al.*, "ISLES 2015 - A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI," *Med. Image Anal.*, vol. 35, pp. 250–269, Jan. 2017, doi: 10.1016/j.media.2016.07.009.

[6] M. R. Hernandez Petzsche *et al.*, "ISLES 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset," *Sci. Data*, vol. 9, no. 1, p. 762, Dec. 2022, doi: 10.1038/s41597-022-01875-5.

[7] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 2018, pp. 3–11. doi: 10.1007/978-3-030-00889-5_1.

[8] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nat. Methods*, vol. 18, no. 2, pp. 203–211, Feb. 2021, doi: 10.1038/s41592-020-01008-z.

[9] M. A. Hambali and P. A. Agwu, "Adversarial Convolutional Neural Network for Predicting Blood Clot Ischemic Stroke," *J. Comput. Theor. Appl.*, vol. 2, no. 1, pp. 51–64, Jun. 2024, doi: 10.62411/jcta.10516.

[10] W. Jiangtao, N. I. R. Ruhaiyem, and F. Panpan, "A Comprehensive Review of U-Net and Its Variants: Advances and Applications in Medical Image Segmentation," *IET Image Process.*, vol. 19, no. 1, Jan. 2025, doi: 10.1049/ipr2.70019.

[11] S. Zuo, Y. Xiao, X. Chang, and X. Wang, "Vision transformers for dense prediction: A survey," *Knowledge-Based Syst.*, vol. 253, p. 109552, Oct. 2022, doi: 10.1016/j.knosys.2022.109552.

[12] C. Asuai *et al.*, "Transformer-Augmented Deep Learning Ensemble for Multi-Modal Neuroimaging-Based Diagnosis of Amyotrophic Lateral Sclerosis," *J. Comput. Theor. Appl.*, vol. 3, no. 2, pp. 190–205, Oct. 2025, doi: 10.62411/jcta.14661.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Lecture Notes in Computer Science*, Springer, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

[14] J. Tagnamas, H. Ramadan, A. Yahyaouy, and H. Tairi, "SCA-InceptionUNeXt: A lightweight Spatial-Channel-Attention-based network for efficient medical image segmentation," *Knowledge-Based Syst.*, vol. 311, p. 113166, Feb. 2025, doi: 10.1016/j.knosys.2025.113166.

[15] C. Szegedy *et al.*, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, pp. 1–9. doi: 10.1109/CVPR.2015.7298594.

[16] A. Rahman *et al.*, "Deep learning-driven segmentation of ischemic stroke lesions using multi-channel MRI," *Biomed. Signal Process. Control*, vol. 105, p. 107676, Jul. 2025, doi: 10.1016/j.bspc.2025.107676.

[17] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas, and K. E. Muller, "Contrast-limited adaptive histogram equalization: speed and effectiveness," in *[1990] Proceedings of the First Conference on Visualization in Biomedical Computing*, pp. 337–345. doi: 10.1109/VBC.1990.109340.

[18] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, "Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks," in *Lecture Notes in Computer Science*, 2017, pp. 379–387. doi: 10.1007/978-3-319-67389-9_44.

[19] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso, "Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations," in *Lecture Notes in Computer Science*, 2017, pp. 240–248. doi: 10.1007/978-3-319-67558-9_28.

[20] G. Kaur and J. Chhaterji, "A Survey on Medical Image Segmentation," *Int. J. Sci. Res.*, vol. 6, no. 5, pp. 1305–1311, 2017, [Online]. Available: https://www.ijsr.net/archive/v6i5/ART20173326.pdf

[21] Y. Zhang, J. M. Gorriz, and Z. Dong, "Deep Learning in Medical Image Analysis," *J. Imaging*, vol. 7, no. 4, p. 74, Apr. 2021, doi: 10.3390/jimaging7040074.
[22] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in *Lecture Notes in Computer Science*, 2018, pp. 3–19. doi: 10.1007/978-3-030-01234-2_1.
[23] Y. Ye *et al.*, "Image segmentation using improved U-Net model and convolutional block attention module based on cardiac magnetic resonance imaging," *J. Radiat. Res. Appl. Sci.*, vol. 17, no. 1, p. 100816, Mar. 2024, doi: 10.1016/j.jrras.2023.100816.
[24] W. Yu, P. Zhou, S. Yan, and X. Wang, "InceptionNeXt: When Inception Meets ConvNeXt," *ArXiv*. Apr. 05, 2025. [Online]. Available: http://arxiv.org/abs/2303.16900
[25] Y. Gao, Y. Jiang, Y. Peng, F. Yuan, X. Zhang, and J. Wang, "Medical Image Segmentation: A Comprehensive Review of Deep Learning-Based Methods," *Tomography*, vol. 11, no. 5, p. 52, Apr. 2025, doi: 10.3390/tomography11050052.
[26] Y. Liu, Z. Zhang, J. Yue, and W. Guo, "SCANeXt: Enhancing 3D medical image segmentation with dual attention network and depth-wise convolution," *Heliyon*, vol. 10, no. 5, p. e26775, Mar. 2024, doi: 10.1016/j.heliyon.2024.e26775.
[27] T. Li *et al.*, "SrSNet: Accurate segmentation of stroke lesions by a two-stage segmentation framework with asymmetry information," *Expert Syst. Appl.*, vol. 254, p. 124329, Nov. 2024, doi: 10.1016/j.eswa.2024.124329.
[28] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv*. Jan. 30, 2017. [Online]. Available: http://arxiv.org/abs/1412.6980