Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 21 No. 9 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v21i09.55677

PAPER

Swin-BSSeg: A Novel Swin Transformer-Enhanced Architecture for Accurate Ischemic Stroke Lesion Segmentation in MRI Images

Prabha Susy Mathew[1](✉), Anitha S Pillai[1], Lazzaro di Biase[2], Ajith Abraham[3]

[1]School of Computing Sciences, Hindustan Institute of Technology and Science, Chennai, Tamil Nadu, India

[2]Neurology Unit, Campus Bio-Medico University of Rome, Rome, Italy

[3]Sai University, Chennai, Tamil Nadu, India

rp.22703019@student.hindustanuniv.ac.in

ABSTRACT

Ischemic stroke, caused by obstructed cerebral blood flow, remains a leading cause of mortality and disability, necessitating precise magnetic resonance imaging (MRI)-based lesion detection. This paper proposes Swin-BSSeg (brain symmetry segmentation) algorithm, an enhanced version of the brain symmetry segmentation network (BSSNet), for improved stroke lesion segmentation. Swin-BSSeg integrates Swin Transformers to capture global context and long-range dependencies through a hierarchical attention mechanism. The encoder replaces traditional convolutions with depth-wise separable convolution (DWSC) blocks comprising three cascaded depth-wise convolutional layers for efficient feature extraction and parameter reduction. Feature transfer to the decoder is accomplished via concatenation operations. The decoder also employs DWSC blocks to reduce computational demands and incorporates attention-guided connections (AGC) to refine the contextual diversity—defined as model's ability to capture varied and spatially distributed lesion features. The model was evaluated on the public datasets (Anatomical Tracings of Lesions after Stroke) and ISLES (Ischemic Stroke Lesion Segmentation), which achieved a Dice Coefficient of 0.842 and an accuracy of 0.87 on the ATLAS dataset, while a Dice Coefficient of 0.8049 and an accuracy of 0.84 on ISLES, outperforming BSSNet in contextual understanding and boundary precision. The improvements on the ATLAS dataset were statistically significant ($p < 0.05$), confirming the reliability of the proposed enhancements.

KEYWORDS

attention guided connection (AGC), ATLAS dataset, BSSNet, ISLES dataset, stroke lesion segmentation, Swin transformers

Mathew, P.S., Pillai, A.S., di Biase, L., Abraham, A. (2025). Swin-BSSeg: A Novel Swin Transformer-Enhanced Architecture for Accurate Ischemic Stroke Lesion Segmentation in MRI Images. *International Journal of Online and Biomedical Engineering (iJOE)*, 21(9), pp. 43–62. https://doi.org/10.3991/ijoe.v21i09.55677

Article submitted 2025-03-23. Revision uploaded 2025-04-27. Final acceptance 2025-05-02.

1 INTRODUCTION

Ischemic stroke remains a formidable global health challenge, ranking as a primary cause of disability and heightened mortality risk. A combination of magnetic resonance imaging (MRI) and advanced segmentation techniques is essential for precise diagnosis and informed treatment strategies.

Brain stroke segmentation employs various strategies, including convolutional neural networks (CNNs), transformers, and hybrid models. Traditional CNNs, widely used for their efficiency, rely on localized filters that restrict their receptive field and capacity to grasp global context and long-range relationships [1]. U-Net and its variants enhance CNNs with skip connections for better localization, whereas fully convolutional networks (FCNs) enable pixel-wise predictions. Transformers overcome CNN limitations by modelling long-range dependencies but require extensive computational resources. Hybrid models combining CNNs and transformers aim to balance efficiency and accuracy.

Efficient and precise segmentation of stroke lesions is critical for evaluating brain damage, because stroke remains a primary contributor to global mortality and disability. Extracting stroke lesion information from MRI scans is a challenging aspect in stroke diagnosis and treatment strategy development. MRI provides full images of the brain tissue, permitting clinicians to consider the level and impact of a stroke. However, manual segmentation of these lesions is time-consuming, less efficient, and susceptible to variability, which drives the research and design of automated approaches to increase their effectiveness. Sathish and Sheet [2], proposed a wrapper method for evaluating the importance of MRI sequences in stroke lesion segmentation using a deep neural network and identified DWI as the optimal sequence for ischemic stroke lesion segmentation.

Recently, deep learning (DL) methods have made significant strides in medical image analysis, particularly in stroke lesion segmentation. CNNs such as CNN-Res [1], 3D CNN [3], modified multi-scale 3D CNN [4], 3D residual CNN [5], multi-path 2.5D CNN [6], and multi-scale region aligned CNN [7] have been extensively utilized owing to their ability to capture spatial hierarchies in imaging data. In particular, the U-Net architecture [8], has achieved significant success in a range of medical image segmentation tasks, such as stroke lesion segmentation. Long-range dependencies and global context are critical in brain lesion segmentation tasks as they enable the model to capture spatial relationships across the entire brain scan, thereby improving accuracy in identifying stroke-affected areas with improved lesion delineation. Convolution-based models, although effective at capturing local features, fall short in ischemic stroke detection due to their limited receptive fields and inability to capture the extensive spatial variability and complexity of stroke lesions, leading to incomplete or inaccurate segmentation.

To overcome these limitations, several transformer-based architectures [9, 10] have been introduced such as TransRender [11], Neuro-TransUNet [12], and hybrid spatio-temporal transformer networks [10], offering improved capabilities for modelling long-range dependencies and global contexts. The integration of Swin transformers is driven by their ability to model complex, long-range dependencies and capture hierarchical spatial information, while maintaining computational efficiency through shifted window operations. This capability is crucial for accurately segmenting ischemic stroke lesions, which are often present as dispersed and irregular morphologies across the brain [13, 9]. By processing images hierarchically and using shifted windows for self-attention computation, the Swin transformer [14] effectively captures complex spatial relationships within the data.

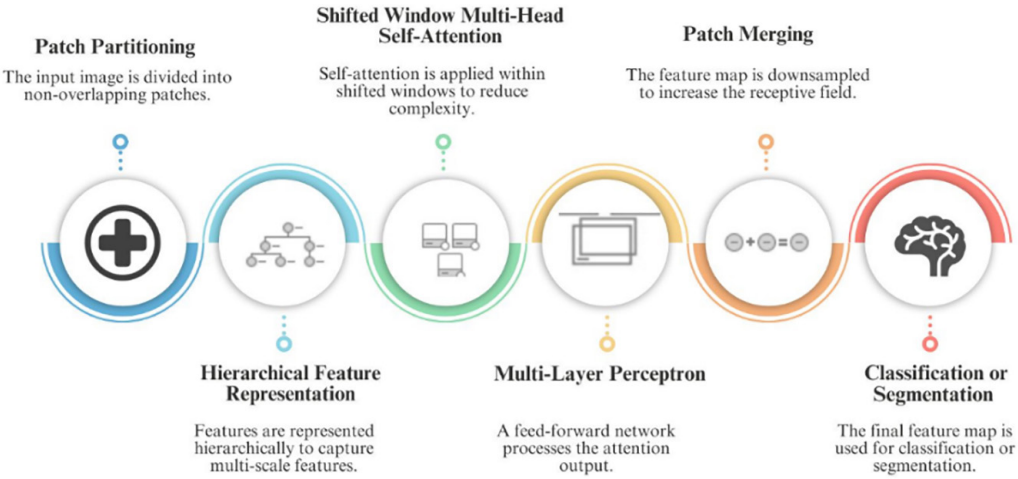

Fig. 1. Swin transformer pipeline

The integration of Swin transformers into U-shaped network architectures, as seen in models such as Swin UNETR [15], has shown promising results in brain tumor segmentation and holds potential for application in stroke lesion segmentation. Figure 1 illustrates the pipeline of a Swin transformer.

This study introduced the Swin-BSSeg, an improved model for stroke lesion segmentation in MRI scans. By integrating the Swin Transformer into the brain symmetry segmentation network (BSSNet), the model enables hierarchical image processing, allowing better analysis of the relationships between different brain regions. This method boosts the segmentation accuracy by capturing both local and global contextual information, making it easier to identify regions in the brain affected by stroke with greater precision. The ability of the model to process information at various scales contributed to its enhanced performance in lesion delineations. Furthermore, the model employs attention-guided connection (AGC) mechanism, which optimizes feature extraction. It integrates low-level features (edges and textures) with high-level semantic features (lesion morphology, location, and brain symmetry patterns). This approach enhances the precision of segmentation while simultaneously reducing the computational demands, resulting in a more efficient model. Compared to recent models such as TransRender and Neuro-TransUNet, Swin-BSSeg introduced a balanced architecture that integrates global feature modeling through Swin Transformers with efficient local feature extraction via DWSC. This design enables improved segmentation performance while maintaining computational efficiency. Although advances in CNN and hybrid models have narrowed performance gaps, there remains a significant gap in achieving high segmentation accuracy with minimal computational overhead while preserving both global and local contextual representations. Swin-BSSeg is proposed to address this gap by leveraging the strengths of Swin Transformers and lightweight convolutional operations.

The structure of this paper is as follows: Section 2 examines related research and developments in segmentation approaches. Section 3 provides a detailed description of the proposed Swin-BSSeg architecture and its associated improvements. Section 4 presents the experimental analysis, evaluation datasets, findings, and performance analysis of the Swin-BSSeg and BSSNet. Finally, Section 5 concludes the paper and summarizes the contributions and potential directions for future research.

2 RELATED WORKS

Precise identification and demarcation of stroke lesions on MRI scans are of paramount importance in stroke diagnosis and treatment planning. MRI technology provides high-resolution brain tissue images that are essential for evaluating stroke severity and outcomes. However, the manual process of lesion segmentation is not only time-consuming and labor-intensive but also susceptible to human error. This underscores the critical need to develop automated methods to enhance both the accuracy and efficiency. Although DL has made significant strides in this field, traditional CNNs often fail to fully capture extensive spatial relationships and overall context. These elements are crucial for accurate identification of lesions that vary in size, shape, and location. This section discusses the latest advancements in neural network architectures and hybrid systems designed to address these challenges.

2.1 Traditional CNN architectures

Convolutional neural networks are instrumental in stroke lesion segmentation. DeepMedic, a 3D CNN model developed by Kamnitsas et al. [16], incorporates conditional random fields and multi-scale pathways, marking one of the earliest successful applications of CNNs in brain lesion segmentation. However, the use of localized filters hinders the capacity of the model to capture global spatial relationships, which are essential for addressing the irregular nature of lesions. Liu et al. [17], introduced a multi-receptive-field CNN that combined information from various spatial scales within an encoder-decoder framework. This innovation improved the segmentation of ischemic stroke lesions by balancing fine details with broader contextual information. Nevertheless, reliance on conventional convolutional operations results in high computational demands, limiting the scalability and efficiency of the model. Alquhayz et al. [18], introduced MCN, a three-phase DL framework that employs a modified residual U-Net for the final segmentation. It was crafted to handle class imbalance and inter-class similarity in the ATLAS dataset, attaining a mean Dice score of 0.754. Yu et al. [19], proposed a Self-Adaptive Normalization Network (SAN-Net), which is a DL model based on U-Net architecture. Their model aimed to improve the adaptability of stroke lesion segmentation across various MRI imaging facilities. To address the differences between sites resulting from diverse scanners, imaging protocols, and patient populations.

2.2 Attention mechanisms and enhanced architectures

Attention mechanisms have emerged as powerful tools for enhancing segmentation performance in medical imaging. Rueckert et al. [20], developed the Attention U-Net, which integrates attention gates to emphasize relevant areas while suppressing unimportant features. This approach significantly improves the segmentation accuracy, particularly in pancreatic segmentation. However, its application to stroke lesion segmentation remains underexplored, and computational overhead continues to be a concern. Li et al. [21], further advanced the field using BSSNet, an architecture that incorporates multi-scale channel attention (MSCA) and depth-wise separable convolutions (DWSC) for stroke lesion segmentation. These enhancements

allow for effective integration of global and local dependencies while reducing computational complexity. The inclusion of AGC further improved contextual information diversity, resulting in more accurate segmentation outcomes. Despite its effectiveness, BSSNet still faces challenges in efficiently capturing the global context, indicating the need for further architectural improvements. Wu et al. proposed FRPNet [22], a sophisticated segmentation network based on the U-Net architecture, incorporating a twin attention gate (TAG) and multi-dimensional attention pooling (MAP) modules. It achieved 60.16% DSC, 36.20 HD, 85.72% DSC, and 27.02 HD on two distinct ischemic stroke datasets.

2.3 Transformer-based approaches

Transformers, originally developed for natural language processing, have been adapted to computer vision tasks because of their ability to model long-range dependencies. Vaswani et al. [23], pioneered this shift using a transformer architecture by introducing self-attention mechanisms to effectively model spatial relationships. Dosovitskiy et al. [24], extended this work to vision tasks using a vision transformer (ViT) that processes images as a sequence of patches. Although ViTs perform well in image classification, they require substantial computational resources and large datasets, making them less practical for medical imaging applications with limited data availability. To address these limitations, Liu et al. [14], introduced the Swin Transformer, which employs a hierarchical structure and shifted windowing scheme to locally compute self-attention while enabling cross-window connections. This innovation significantly reduces the computational overhead while maintaining high performance across various computer vision tasks. While Swin Transformers have shown promise in medical imaging, their potential in stroke lesion segmentation has not yet been fully explored.

2.4 Hybrid models

Hybrid architectures that combine CNNs and transformers have attracted attention because of their ability to leverage the strengths of both approaches. Malik et al. [25], in their study offered a comprehensive analysis of DL architectures for stroke lesion segmentation, examining CNN-based and transformer-based models and exploring their application across three major public datasets (ATLAS, ISLES and AISD). A prime example is UNETR, created by Hatamizadeh et al. [15], which incorporates a transformer-based encoder and CNN-based decoder in a U-shaped configuration. This structure effectively captures both global and local spatial information, resulting in an exceptional performance in the segmentation of volumetric medical images. Despite its success, the potential application of UNETR in analyzing stroke lesions remains largely unexplored, presenting an opportunity for future research.

2.5 Challenges and research gaps

Although advancements have been made, segmentation of stroke lesions remains challenging. Conventional CNNs are adept at extracting local features but struggle

with modelling global relationships. Although transformer-based models address this limitation, they introduce computational complexity and require large datasets. Hybrid models offer a promising alternative however, their development for this specific application is still in its early stages. This study contributes to the field by proposing the Swin-BSSeg algorithm, which combines the hierarchical capabilities of Swin transformers with the computational efficiency of DWSC and the contextual richness of AGC. This innovative approach addresses existing challenges by improving feature representation, reducing computational requirements, and enhancing the segmentation accuracy of stroke lesion images.

3 METHOD

This study introduced an automated framework for detecting and quantifying stroke lesions on MRI. The study employs DL based segmentation models, specifically CNNs, U-Net, and Transformer-based architectures such as the Swin transformer to segment stroke lesions, utilizing the ATLAS and ISLES datasets for evaluation. This methodology uses MRI scans and lesion masks, that are essential for supervised learning. The MRI Lesion segmentation process is illustrated in Figure 2. These images underwent a series of preprocessing steps to enhance the uniformity and eliminate extraneous elements. The process includes intensity normalization for consistent contrast, skull stripping to isolate brain regions, and cropping to a specific region of interest (ROI) to reduce background interference. The images were then resized to 192 × 192 pixels for uniform input dimensions and subjected to Z-score normalization to standardize the intensity values across the datasets. Finally, the pre-processed images and masks are converted into NumPy (.npy) for efficient storage and rapid access during training and inference.

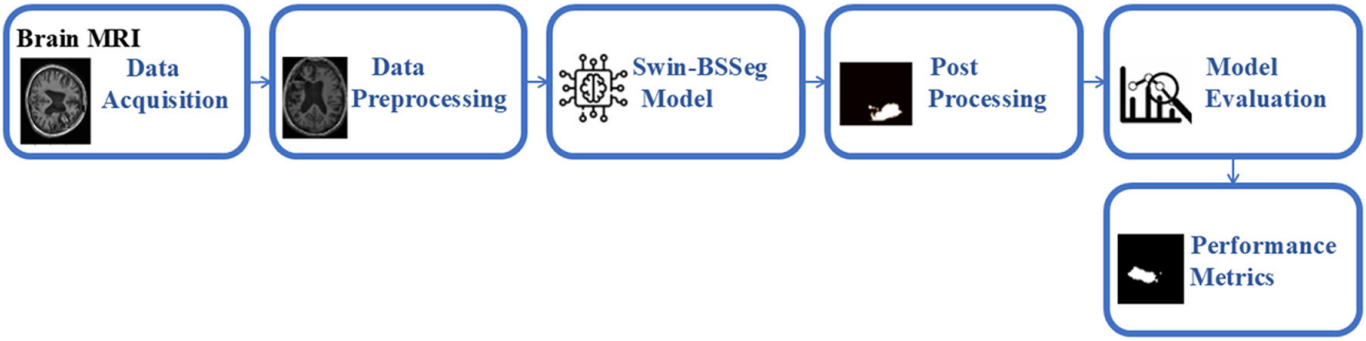

Fig. 2. Steps in MRI lesion segmentation

The core of the system comprises a hybrid transformer-CNN based architecture that integrates convolutional and attention-based methods to enhance segmentation accuracy. It incorporates DWSC for efficient feature extraction, Swin Transformer Blocks for processing MRI slices using multi-head self-attention and hierarchical feature extraction, and MSCA to refine the extracted features by highlighting important lesion areas. This combination enables the model to capture both the local and global structural details of stroke lesions.

Following the model processing, post-processing techniques were applied to refine the predicted masks and improve their accuracy. These include thresholding to convert the probability outputs into binary segmentation masks. The lesion

volume was estimated by multiplying the total number of segmented pixels, which were then converted to cubic millimeters (mm^3) and cubic centimeters (cm^3) for clinical relevance. The model performance was evaluated by comparing its output to expert-annotated masks using metrics such as the dice coefficient, intersection over union (IoU), precision, and recall.

This study introduces the Swin-BSSeg algorithm, which incorporates a Swin transformer (SWIN) to enhance the precision of skin lesion segmentation. The integration of the SWIN enables the capture of broad contextual information and extensive interconnections.

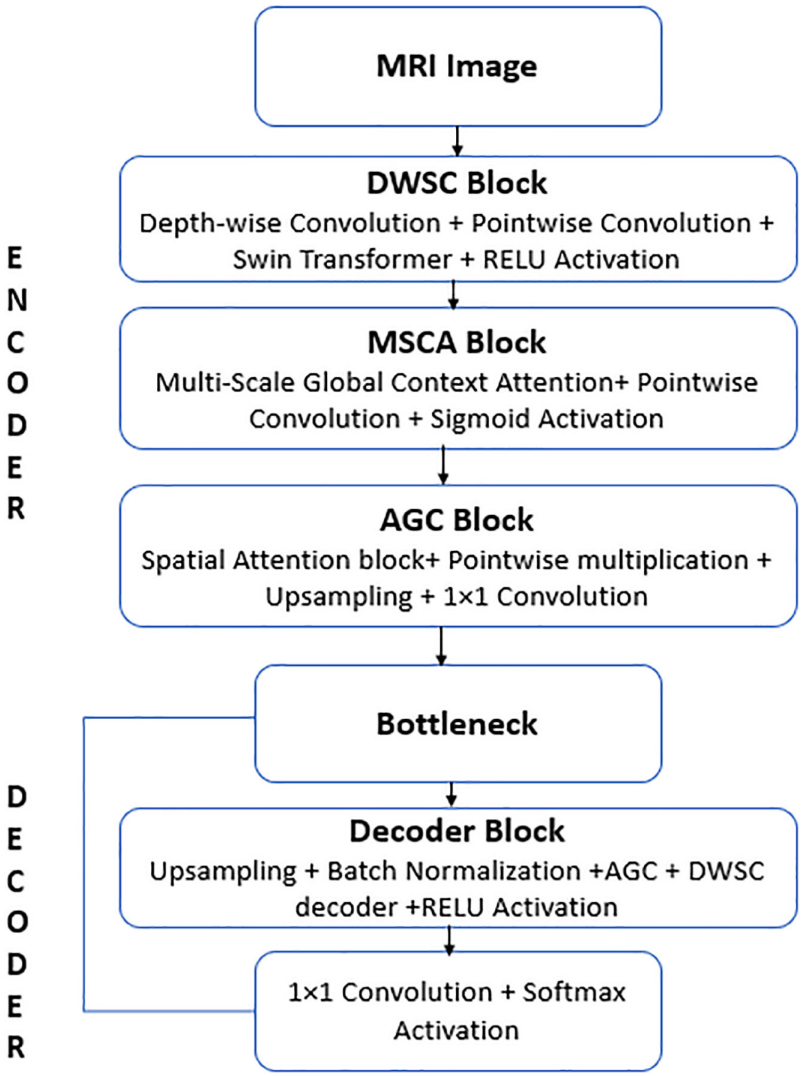

Fig. 3. The core architecture of the proposed system

This architecture processes images using a hierarchical approach, thereby enhancing the capacity of the models to analyze relationships across various image regions. It has also been applied to the management of various complex lesions. By merging DWSC blocks and AGC, the Swin-BSSeg model provides the best feature extraction and increased feature representation, thereby improving segmentation performance. Figure 3 illustrates the architecture of the proposed system.

Algorithm 1: Swin-BSSeg Algorithm for Stroke Lesion Segmentation

Input: 3D brain MRI volume (T1/T2/FLAIR/DWI), 256×256×Z voxels, normalized [0,1]
Step 1: Feature Extraction using DWSC Block
Procedure Swin-BSSeg(input_image)
encoded_features, skip_connections ← Encoder(input_image)
segmentation_mask ← Decoder (encoded_features, skip_connections)
return segmentation_mask
Procedure DWSC_Block (input, output_channels)
depth_wise ← DepthWiseConv2D (3×3, padding='same') (input)
norm1 ← BatchNormalization () (depth_wise)
act1 ← ReLU () (norm1)
point_wise ← Conv2D (output_channels, 1×1) (act1)
output ← ReLU (BatchNormalization () (point_wise))
return output
Procedure Cascaded_DWSC_Block (input, channels)
x ← DWSC_Block (input, channels)
x ← DWSC_Block (x, channels)
x ← DWSC_Block (x, channels)
return x
Step 2: Global Context Modeling using Swin Transformer
Procedure SwinTransformerBlock (input, dim, heads, window_size, shift_size=0)
norm1 ← LayerNorm () (input)
if shift_size > 0 then
    shifted_x ← CyclicShift(shift_size) (norm1)
    windows ← WindowPartition (shifted_x, window_size)
else
    windows ← WindowPartition (norm1, window_size)
attn_windows ← WindowAttention (windows, dim, heads, qkv_bias=True)
x ← WindowReverse (attn_windows, window_size, shift= (shift_size > 0), shift_size=shift_size)
x ← input + x
norm2 ← LayerNorm () (x)
output ← x + MLP (norm2)
return output
Step 3: Feature Refinement using Multi-Scale Channel Attention (MSCA)
Procedure MSCA_Module(input)
avg_pool ← GlobalAveragePooling2D(keepdims=True) (input)
scale1 ← Conv2D (input_channels, 1×1) (avg_pool)
scale3 ← Conv2D (input_channels, 3×3, padding='same') (avg_pool)
scale5 ← Conv2D (input_channels, 5×5, padding='same') (avg_pool)
concat ← Concatenate ([scale1, scale3, scale5])
channel_weights ← Sigmoid (Conv2D (input_channels, 1×1) (concat))
output ← Add ([input, Multiply ([input, channel_weights])])
return output
Step 4: Context-Aware Upsampling using Attention-Guided Connections (AGC)
Procedure AGC_Module (low_features, high_features)
if shape(high_features) ≠ shape(low_features) then
    high_features ← UpSampling2D () (high_features)
end if
attn_map ← Sigmoid (BatchNormalization (Conv2D (1×1) (high_features)))
attended ← Multiply ([low_features, attn_map])
merged ← Concatenate ([attended, high_features])
output ← ReLU (BatchNormalization (Conv2D (3×3, padding='same') (merged)))
return output

(Continued)

Algorithm 1: Swin-BSSeg Algorithm for Stroke Lesion Segmentation (Continued)

Step 5: Encoder and Decoder Design
Procedure Encoder(input_image)
x ← ReLU (BatchNormalization (Conv2D (64, 3×3, padding='same') (input_image)))
skip_connections ← []
channels ← [64, 128, 256, 512]
for i = 1 to 4 do
    x ← Cascaded_DWSC_Block (x, channels[i])
  append x to skip_connections
    x ← MaxPooling2D(pool_size=2) (x)
end for
x ← SwinTransformerBlock (x, dim=512, heads=8, window_size=7)
x ← SwinTransformerBlock (x, dim=512, heads=8, window_size=7, shift_size=3)
x ← MSCA_Module(x)
return x, skip_connections
Procedure Decoder (bottleneck, skip_connections)
x ← bottleneck
channels ← [256, 128, 64, 32]
for i = 1 to 4 do
    x ← UpSampling2D(size=2) (x)
    x ← AGC_Module(skip_connections[3-i], x)
    x ← Cascaded_DWSC_Block (x, channels[i])
end for
x ← MSCA_Module(x)
return output ← Conv2D (1, 1 × 1, activation='sigmoid') (x)
Step 6: Output Segmentation Mask Generation
Generate and return the final binary lesion mask from the decoder output.

The model incorporates an encoder to optimize feature extraction while simultaneously reducing computational complexity. The key innovation lies in the integration of SWIN transformers into DWSC, which considerably increases the accuracy of the model compared to conventional convolutional layers. *DWSC* splits the convolution procedure into two functions: depth-wise and point-wise convolution. It has been used to study complex spatial features while reducing the computation time by minimizing parameters. In the depth-wise convolution stage, each input channel was processed separately using a convolutional filter, focusing on spatial filtering within the channel. The output from each channel is subsequently merged using a 1 × 1 pointwise convolution to integrate channel-wise information. This can be expressed using the following equation:

$$DWSC(x) = \sigma(W_p * (W_d * x)) \quad (1)$$

In this context, W_d represents the depth-wise convolution filter, W_p represents the point-wise convolution filter, and σ denotes the ReLU activation function. The *DWSC* technique reduces the memory consumption and computational complexity through these two processes. Consequently, the number of required parameters was reduced.

To promote effective gradient flow and lower the possibility of vanishing gradients during training, residual connections were integrated into the encoder. These connections help preserve important feature information by allowing gradients to pass directly through network layers. This can be mathematically represented as follows:

$$Residual(x) = x + DWSC \quad (2)$$

This formula ensures that the original input x is added back to the *DWSC* output, helping retain the essential features.

The proposed architecture, as depicted in Figure 3, leverages Swin Transformers [14], in the feature integration stage to effectively capture global context and long-range dependencies. This approach enhances the model's ability to comprehend the spatial relationships within the input images. Swin Transformers function by dividing the input image into discrete, non-overlapping patches, each of which undergoes independent processing. This integration significantly improves the capacity of the model to analyze and interpret complex visual data across various scales and dimensions.

The Swin transformer model employs a patch-based approach that balances computational efficiency with the ability to capture complex spatial relationships. By processing the input image in non-overlapping segments, the model applies a series of transformations through its layers, progressively merging the features to construct hierarchical representations. This process can be summarized as follows:

$$Swin(x) = Merge(Patch(x)) \tag{3}$$

In this formulation, $Patch(x)$ represents the initial segmentation of the input image into discrete patches, whereas Merge denotes the subsequent integration of these patch-level features. This hierarchical merging procedure enables the model to synthesize a comprehensive representation of the input, effectively capturing both local and global spatial dependencies within the image. The self-attention mechanism evaluates the significance of different areas by computing the attention scores among the patches. This process highlighted the crucial components of each patch. The mathematical formulation of the self-attention mechanism [8] is as follows:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{4}$$

where Q, K, V denote the query, key, and value matrices respectively; and d_k denotes the dimensions of the key.

The encoder transfers the extracted features to the decoder via concatenation, thereby facilitating seamless integration. This process enables the decoder to accurately reproduce the segmented output. By combining features from multiple layers through concatenation, the decoder receives a diverse and comprehensive set of features, which is essential for precise segmentation. This mechanism can be expressed as:

$$Concat(x_1, x_2) = [x_1, x_2] \tag{5}$$

The feature maps x_1 and x_2 are merged through a concatenation function, enabling the decoder to access a broad spectrum of information for segmentation purposes. To maintain simplicity and efficiency, the decoder utilizes depthwise separable convolution (DWSC) blocks. These blocks facilitate the creation of high-resolution segmented outputs without unnecessary computational complexity. Conventional connection operations are replaced by an *AGC* to improve the capacity of the model to concentrate on crucial contextual information. The hierarchical attention structure enables the network to extract and merge features at multiple levels of abstraction. This layered representation helps distinguish subtle lesion characteristics while retaining broader spatial coherence, thereby improving segmentation accuracy across varied lesion shapes and sizes. Within *AGC*, attention scores are employed to emphasize important features and focus on the most relevant areas of the input image. This process is mathematically expressed as follows:

$$AGC\ (x) = Attention(x_{low}, x_{high}) \tag{6}$$

The Swin-BSSeg architecture employs cutting-edge methods to improve the feature extraction, incorporate the global context, and streamline the reconstruction of the segmented results. Through the synergistic combination of DWSC, Swin Transformers, and AGC techniques, this model achieves substantial enhancements in its capacity to segment lesions from MRI stroke images. Overfitting was mitigated through dropout layers, early stopping, and extensive data augmentation, including rotation, flipping, and intensity variation. Consequently, Swin-BSSeg offers a powerful alternative to traditional approaches, effectively addressing the limitations of conventional segmentation methods.

The Swin-BSSeg architecture is designed with computational efficiency in mind. By integrating DWSC, the model significantly reduces the number of trainable parameters and floating-point operations (FLOPs) compared to traditional convolutional models. The use of Swin Transformers with a shifted window mechanism further minimizes the self-attention computation by limiting it to local non-overlapping windows, while still enabling cross-window information flow. This hierarchical design ensures that global context is captured without incurring the high computational cost typical of full self-attention. The combination of DWSC and window-based attention leads to a lightweight yet powerful model that achieves competitive accuracy with lower memory usage and faster inference time.

3.1 Experimental setup and performance analysis

A computer with an NVIDIA GeForce RTX 3050 GPU and an Intel Core i7 12th generation processor was used for this study. This hardware setup was used for training, validation, and testing to guarantee an effective computation and faster model performance. While the high-performance CPU ensured that data pretreatment and administration proceeded smoothly, the GPU's parallel processing capabilities allowed DL activities to be completed more quickly.

ATLAS dataset. The ATLAS dataset [26], functions as an open-access repository for brain MRI scans and is specifically curated for the evaluation of lesion segmentation methods. The ATLAS dataset, comprises T1-weighted MRI scans. These images are shown in Figure 4. This collection encompasses high-resolution MRI data acquired from a diverse group of patients exhibiting various stroke-induced lesions. Each image within the dataset was complemented by precise ground-truth lesion masks to establish a uniform standard for assessing the accuracy of segmentation techniques.

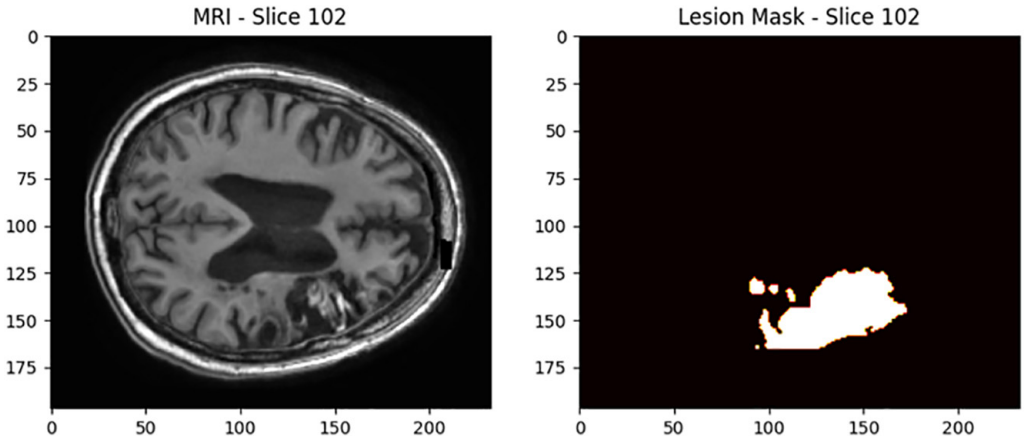

Fig. 4. MRI slice with lesion mask in ATLAS dataset

ISLES dataset. The ISLES dataset [27], serves as an open-access resource for assessing stroke lesion image segmentation techniques. It comprises MRI scans from patients with ischemic stroke, which are complete with detailed lesion annotations. Images from the ISLES dataset are shown in Figure 5. The ISLES dataset contains 400 multi-vendor MRI scans with FLAIR, DWI, T1, and T2-weighted scans. This collection is particularly useful because of its emphasis on complex stroke lesions, which provides a wide array of lesion types and cases that pose significant segmentation challenges. The research community highly valued this dataset for its intricacy and variety.

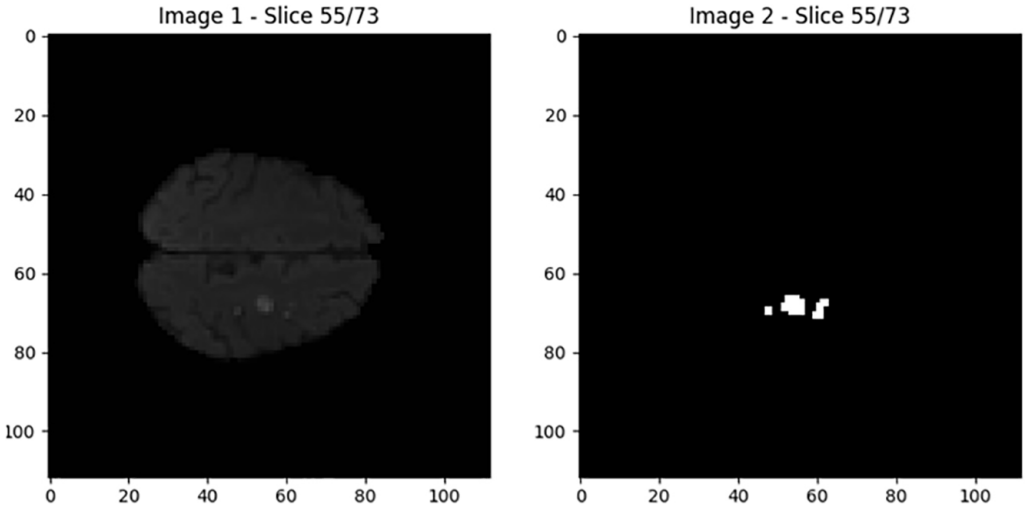

Fig. 5. MRI slice with lesion mask in ISLES dataset

Training and testing pipelines. Of the 955 available MRI scans for the ATLAS dataset, 655 scans were used for training, 150 for validation, and 150 for testing. The ISLES22 dataset comprised 250 multimodal MRI images, of which 150 scans were used for training, 50 for validation, and 50 for testing, ensuring a balanced evaluation of the model performance. The model was trained for 50 epochs and a learning rate of 0.0001, using carefully chosen training and validation sets to maintain a wide variety of lesions and to ensure robust generalization. The loss function employed was Dice Loss, which is particularly well-suited for segmenting medical images because it can handle class imbalances and enhances the precision of lesion boundary identification.

3.2 Evaluation metrics

The following metrics were used to evaluate the segmentation models [28]:

Dice Coefficient (Dice): Dice was used to evaluate the difference between the ground truth and detected lesion masks. This is computed as follows:

$$Dice = \frac{2TP}{2TP + FN + FP} \tag{7}$$

where: *TP* is True Positives (Correctly detects a condition that is present)

TN is True Negatives (Correctly detects a condition that is absent)

FP is False Positives (Incorrectly detects a condition that is not present)

FN is False Negatives (Fails to detect a condition that is present)

Intersection Over Union (IoU): IoU (Intersection over Union) is a metric that evaluates the degree of overlap between detected and ground truth lesions, relative to the total area they encompass. This was calculated using the following formula:

$$IoU = \frac{TP}{TP + FP + FN} \tag{8}$$

An increased IoU range demonstrates a more precise segmentation.

Precision: Precision is used to calculate the fraction of true positive pixels among all the forecasted positive pixels:

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

Recall: Recall is used to measure the fraction of true positive pixels among all actual positive pixels:

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

Higher values of these metrics indicate improved segmentation performance.

4 RESULTS AND DISCUSSION

The BSSNet model and Swin-BSSeg were evaluated on both ATLAS and ISLES datasets. The results of these methods with their evaluation metrics are presented in Tables 1 and 2.

Table 1. Performance comparison on ATLAS Dataset

Model\Metric	ATLAS Dataset					
	Dice Coefficient	IoU	Precision	Recall	F1-Score	Accuracy
BSSNet	0.80	0.8055	0.7980	0.8199	0.8055	0.82
Swin-BSSeg	0.842	0.8443	0.8502	0.8699	0.8580	0.87

Table 2. Performance comparison on ISLES Dataset

Model\Metric	ISLES Dataset					
	Dice Coefficient	IoU	Precision	Recall	F1-Score	Accuracy
BSSNet	0.7309	0.7459	0.77	0.79	0.7846	0.80
Swin-BSSeg	0.8049	0.7947	0.8186	0.8401	0.8265	0.84

The experimental results presented in Tables 1 and 2, and Figure 6 to 11 provide a comparative analysis of the BSSNet and Swin-BSSeg models on two prominent datasets: ATLAS and ISLES.

On the ATLAS dataset, the Swin-BSSeg model gave the highest Dice Coefficient of 0.842, which is slightly higher than BSSNet's 0.80, indicating a marginal improvement in the overlap between the predicted and actual lesion masks. However, Swin-BSSeg exhibited a higher IoU of 0.8443 than that of BSSNet (0.8055).

On the ISLES dataset, the Swin-BSSeg model provided an increased performance. A Dice Coefficient of 0.8049 compared to BSSNet's 0.7309, indicates a notable improvement in segmentation accuracy. Similarly, the IoU for Swin-BSSeg was 0.7947, which was higher than that of BSSNet (0.7459), reflecting its enhanced ability to capture lesion regions accurately.

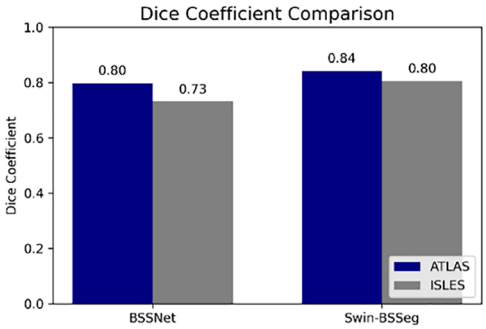

Fig. 6. Dice coefficient

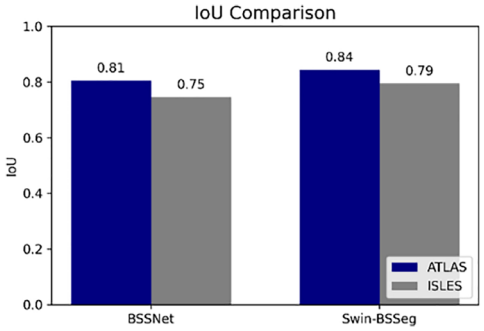

Fig. 7. Intersection over union

In terms of Precision and Recall, Swin-BSSeg achieved scores of 0.8502 and 0.8699, respectively, outperforming BSSNet by 0.7980 and 0.8199 respectively on the ATLAS dataset. Precision and Recall scores for Swin-BSSeg were 0.8186 and 0.8401 was obtained respectively, compared to BSSNet's 0.77 and 0.79, showing improved identification of true lesion pixels while reducing false positives on ISLES dataset.

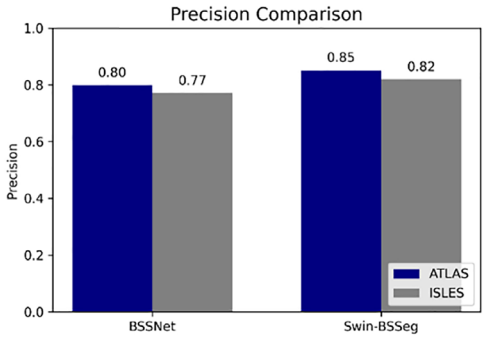

Fig. 8. Precision

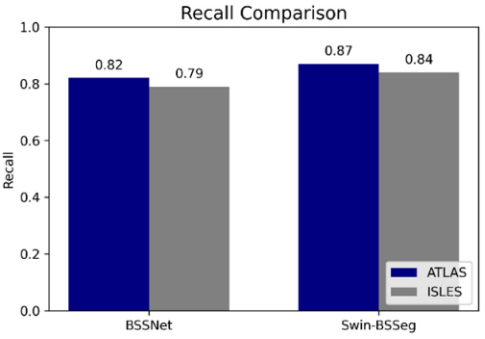

Fig. 9. Recall

On the ATLAS dataset, Swin-BSSeg also showed an F1-score of 0.8580, surpassing BSSNet by 0.8055, demonstrating its superior ability to balance precision and recall. The overall accuracy of Swin-BSSeg was 0.87, indicating an improvement over that of BSSNet (0.82). Experiments with varying AGC attention map sizes and channel depths showed that increasing attention resolution marginally improved Dice by ~1.2%, especially on lesions near cortical folds. However, deeper attention layers led to diminishing returns and higher computational cost, indicating that a balanced AGC configuration offers the best trade-off. Swin-BSSeg gives an increased F1-score of 0.8265 vs. BSSNet gives 0.7846 and an accuracy of 0.84 for Swin-BSSeg when compared to 0.80 for BSSNet on the ISLES dataset.

Further analysis revealed that Swin-BSSeg's segmentation accuracy is influenced by lesion size. The model performs well on large and medium-sized lesions, but it's Dice score decreases significantly for lesions smaller than 5 cm^3. This is likely due to the subtle intensity differences and reduced context in small lesions. Future work could address this limitation using size-aware training strategies or incorporating

focal loss to better emphasize under-represented small lesion cases. Swin-BSSeg handles small or irregularly shaped lesions using the combined power of hierarchical attention and MSCA. These modules enhance sensitivity to shape deformation and faint textures. Nonetheless, performance may drop on extremely small lesions with low contrast, where boundary definition becomes ambiguous.

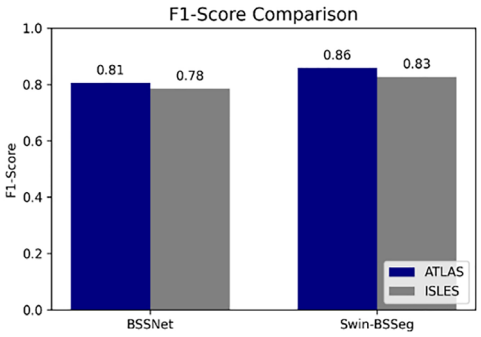

Fig. 10. F1-score

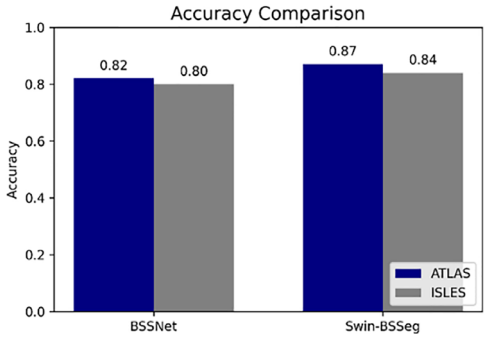

Fig. 11. Accuracy

Figures 12 and 13 show the actual and predicted overlays of the sample images from ATLAS and ISLES.

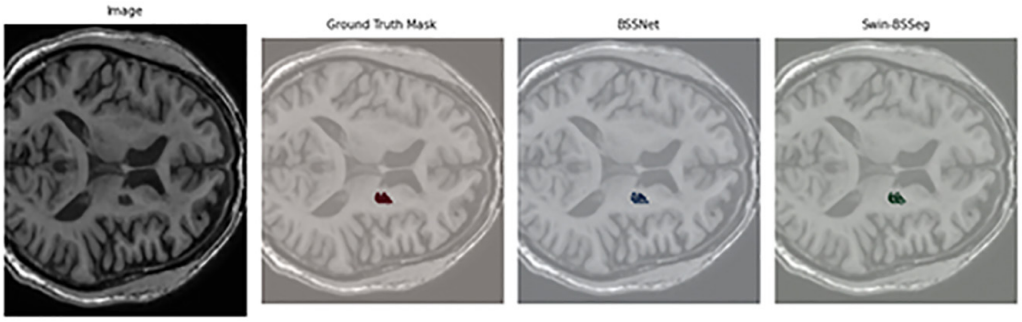

Fig. 12. Actual and prediction overlay on sample image (ATLAS Dataset)

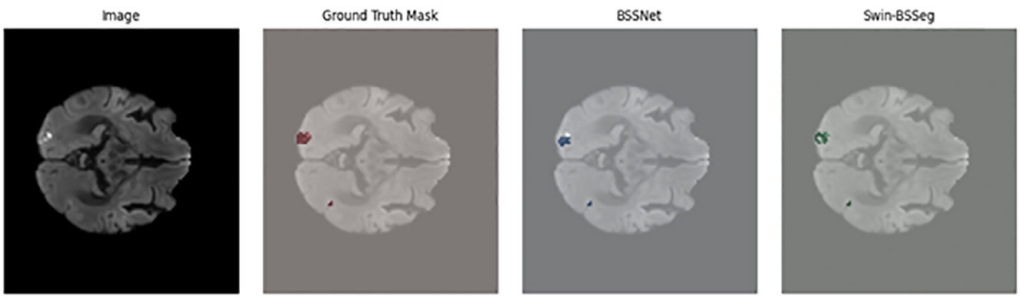

Fig. 13. Actual and prediction overlay on sample image (ISLES Dataset)

The Wilcoxon signed-rank test [13] was used to compare the performances of Swin-BSSeg and BSSNet on the ATLAS and ISLES datasets. Table 3 presents the results of the analysis. Analysis of the ATLAS dataset revealed a p-value of 0.0339, which is below the conventional threshold of 0.05. This indicates that the improved performance of the Swin-BSSeg is statistically significant, suggesting a meaningful enhancement in the segmentation accuracy of ATLAS. Conversely, the ISLES dataset analysis yielded a p-value of 0.0899, exceeding the threshold of 0.05.

Table 3. Wilcoxon signed-rank test of the dice coefficients obtained from the Swin-BSSeg model and the dice scores obtained from BSSNet

Datasets	P Value
ATLAS	0.03394458
ISLES	0.0898562

This implies that the observed difference lacks statistical significance, and the performance disparity between Swin-BSSeg and BSSNet for ISLES may be attributable to chance rather than genuine improvement. These findings demonstrate that while Swin-BSSeg shows a significant advantage when applied to the ATLAS dataset, its superiority over BSSNet is less evident for ISLES. This discrepancy highlights the potential dataset-dependent nature of the effectiveness of Swin-BSSeg.

Evaluation metrics across both ATLAS and ISLES datasets revealed that Swin-BSSeg outperformed BSSNet in terms of overall effectiveness. The incorporation of Swin Transformers into the BSSNet framework enhances the capacity of the model to comprehend the global context, resulting in a markedly improved segmentation accuracy. These findings underscore the greater adaptability and reliability of the Swin-BSSeg for tasks involving lesion segmentation. The Swin-BSSeg model performed better on the ATLAS dataset compared to ISLES due to differences in modality complexity, lesion heterogeneity, and annotation consistency. ATLAS contains T1-weighted images with more consistent contrast and cleaner boundaries, while ISLES includes multimodal images with diverse acquisition settings and more complex lesion appearances, increasing segmentation difficulty.

5 CONCLUSION

This study proposed Swin-BSSeg, a novel stroke lesion segmentation architecture that enhances the traditional BSSNet framework by incorporating Swin transformers, DWSC, MSCA, and AGC. These components work synergistically to capture both local and global contextual features, leading to improved delineation of ischemic stroke lesions from MRI images. The model achieved significant performance improvements over the baseline BSSNet, as demonstrated by higher dice coefficients, precision, recall, and accuracy on the ATLAS and ISLES datasets. Particularly, Swin-BSSeg excelled in handling lesions with complex shapes and varying sizes, showcasing its robustness and clinical relevance.

Beyond segmentation accuracy, the model also demonstrated computational efficiency, with an inference time of approximately 40 ms per image on an RTX 3050 GPU, suggesting its feasibility for near real-time clinical deployment. Qualitative results further affirmed the model's capacity to produce precise segmentation maps with minimal false positives or negatives.

Future work can focus on extending Swin-BSSeg's capabilities in several directions. Integrating multi-modal data—such as CT, FLAIR, and perfusion imaging—with MRI could enrich the input information and enhance segmentation in regions with low contrast. Additionally, adapting the model for real-world clinical applications will require addressing challenges such as imaging protocol variations, scanner differences, and real-time interpretability. Given its lightweight architecture, Swin-BSSeg holds potential for deployment on edge devices, including portable diagnostic tools for emergency stroke assessment.

Another promising direction is to extend the framework to other neurological conditions, such as multiple sclerosis, brain tumors, or traumatic brain injuries, by retraining on task-specific datasets. Moreover, implementing self-supervised or semi-supervised learning strategies may reduce the reliance on large annotated datasets, improving generalization across diverse clinical environments while preserving accuracy.

In conclusion, Swin-BSSeg provides a powerful, efficient, and scalable solution for ischemic stroke lesion segmentation, offering a solid foundation for future advancements in automated neuroimaging diagnostics.

6 DATA AVAILABILITY

The data that support the findings of this study are openly available [ALTAS R2.0] at https://fcon_1000.projects.nitrc.org/indi/retro/atlas.html [26], and [ISLES 2022] at https://zenodo.org/records/7960856 [27].

7 REFERENCES

[1] Y. Gheibi, K. Shirini, S. N. Razavi, M. Farhoudi, and T. Samad-Soltani, "CNN-Res: Deep learning framework for segmentation of acute ischemic stroke lesions on multimodal MRI images," *BMC Med. Inform. Decis. Mak.*, vol. 23, 2023. https://doi.org/10.1186/s12911-023-02289-y

[2] R. Sathish and D. Sheet, "A wrapper method for finding optimal subset of multimodal magnetic resonance imaging sequences for ischemic stroke lesion segmentation," *Comput. Biol. Med.*, vol. 185, 2025. https://doi.org/10.1016/j.compbiomed.2024.109590

[3] A. Dobshik *et al.*, "Acute ischemic stroke lesion segmentation in non-contrast CT images using 3D convolutional neural networks," *arXiv preprint arXiv:2301.06793*, 2023. https://doi.org/10.48550/arxiv.2301.06793

[4] E. Ruthra and A. R. Bevi, "A deep learning based Ischemic stroke lesion segmentation using modified multi-scale 3D CNN," in *Proc. RMKMATE*, 2023, pp. 1–4. https://doi.org/10.1109/RMKMATE59243.2023.10368691

[5] N. Tomita, M. E. Maeder, S. Jiang, and S. Hassanpour, "Automatic post-stroke lesion segmentation on MR images using 3D residual convolutional neural network," *NeuroImage Clin.*, vol. 27, 2020. https://doi.org/10.1016/j.nicl.2020.102276

[6] Y. Xue *et al.*, "A multi-path 2.5 dimensional convolutional neural network system for segmenting stroke lesions in brain MRI images," *arXiv preprint arXiv:1905.10835*, 2019. https://doi.org/10.48550/arxiv.1905.10835

[7] R. Karthik, R. Menaka, M. Hariharan, and D. Won, "Ischemic lesion segmentation using ensemble of multi-scale region aligned CNN," *Comput. Methods Programs Biomed.*, vol. 200, 2021. https://doi.org/10.1016/j.cmpb.2020.105831

[8] O. Ronneberger, "Invited talk: U-Net convolutional networks for biomedical image segmentation," in *Bildverarbeitung für die Medizin 2017*, Maier-Hein, K. Fritzsche, T. L. Deserno, H. Handles, and T. Tolxdorff, Eds., Berlin Heidelberg, Springer, Charm, 2017, p. 3. https://doi.org/10.1007/978-3-662-54345-0.3

[9] T. Anjali, S. Abhishek, and S. Remya, "A novel framework for rosacea detection using Swin transformers and explainable artificial intelligence," *Alexandria Eng. J.*, vol. 118, pp. 36–58, 2025. https://doi.org/10.1016/j.aej.2024.12.080

[10] K. Amador *et al.*, "Hybrid spatio-temporal transformer network for predicting ischemic stroke lesion outcomes from 4D CT perfusion imaging," in *Medical Image Computing and Computer Assisted Intervention*, L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, and S. Li, Eds., Springer, Charm, 2022, pp. 644–654. https://doi.org/10.1007/978-3-031-16437-8_62

[11] Z. Wu, J. Li, X. Zhang, S. Wang, and F. Li, "TransRender: A transformer-based boundary rendering segmentation network for stroke lesions," *Front. Neurosci.*, vol. 17, 2023. https://doi.org/10.3389/fnins.2023.1259677

[12] M. Nouman, M. Mabrok, and E. Rashed, "Neuro-TransUNet: Segmentation of stroke lesion in MRI using transformers," *arXiv preprint arXiv:2406.06017*, 2024. https://doi.org/10.48550/arxiv.2406.06017

[13] Y. Cai *et al.*, "Swin Unet3D: A three-dimensional medical image segmentation network combining vision transformer and convolution," *BMC Med. Inform. Decis. Mak.*, vol. 23, 2023. https://doi.org/10.1186/s12911-023-02129-z

[14] Z. Liu *et al.*, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. ICCV*, 2021, pp. 9992–10002. https://doi.org/10.1109/ICCV48922.2021.00986

[15] A. Hatamizadeh, D. Yang, H. Roth, and D. Xu, "UNETR: Transformers for 3D medical image segmentation," *arXiv preprint arXiv:2103.10504*, 2021. https://doi.org/10.48550/arxiv.2103.10504

[16] K. Kamnitsas *et al.*, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, 2017. https://doi.org/10.1016/j.media.2016.10.004

[17] L. Liu, F.-X. Wu, J. Wang, and Y.-P. Wang, "Multi-receptive-field CNN for semantic segmentation of medical images," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 11, pp. 3215–3225, 2020. https://doi.org/10.1109/JBHI.2020.3016306

[18] H. Alquhayz, H. Z. Tufail, and B. Raza, "The multi-level classification network (MCN) with modified residual U-Net for ischemic stroke lesions segmentation from ATLAS," *Comput. Biol. Med.*, vol. 151, 2022. https://doi.org/10.1016/j.compbiomed.2022.106332

[19] W. Yu *et al.*, "SAN-Net: Learning generalization to unseen sites for stroke lesion segmentation with self-adaptive normalization," *Comput. Biol. Med.*, vol. 156, 2023. https://doi.org/10.1016/j.compbiomed.2023.106717

[20] D. Rueckert *et al.*, "Attention U-Net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018. https://doi.org/10.48550/arxiv.1804.03999

[21] Z. Li *et al.*, "A novel multi-scale channel attention-guided neural network for brain stroke lesion segmentation," *IEEE Access*, vol. 11, pp. 66050–66062, 2023. https://doi.org/10.1109/ACCESS.2023.3289909

[22] Z. Wu, X. Zhang, F. Li, S. Wang, and J. Li, "A feature-enhanced network for stroke lesion segmentation from brain MRI images," *Comput. Biol. Med.*, vol. 174, 2024. https://doi.org/10.1016/j.compbiomed.2024.108326

[23] A. Vaswani *et al.*, "Attention is all you need," *arXiv preprint arXiv:1706.03762*, 2017. https://doi.org/10.48550/arxiv.1706.03762

[24] A. Kolesnikov *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020. https://doi.org/10.48550/arxiv.2010.11929

[25] M. Malik *et al.*, "Stroke lesion segmentation and deep learning: A comprehensive review," *Bioengineering*, vol. 11, no. 1, p. 86, 2024. https://doi.org/10.3390/bioengineering11010086

[26] S. Liew *et al.*, "A large, curated, open-source stroke neuroimaging dataset to improve lesion segmentation algorithms," *bioRxiv preprint*, 2021. https://doi.org/10.1101/2021.12.09.21267554

[27] M. Petzsche *et al.*, "ISLES 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset," *Sci. Data*, vol. 9, 2022. https://doi.org/10.1038/s41597-022-01875-5

[28] R. Ahmed, A. Al Shehhi, N. Werghi, and M. L. Seghier, "Segmentation of stroke lesions using transformers-augmented MRI analysis," *Hum. Brain Mapp.*, vol. 45, no. 11, p. e26803, 2024. https://doi.org/10.1002/hbm.26803

8 AUTHORS

Ms. Prabha Susy Mathew is currently pursuing her doctoral studies at Hindustan Institute of Technology and Science (HITS), Chennai. She has over 14 years of academic experience, along with prior industry experience as a Test Engineer. Her research interests include Artificial Intelligence, Machine Learning, Data Mining, Big Data, and the Internet of Things (IoT). She has authored and co-authored over 14 publications, including journal articles, international conference papers, and book chapters. Her work has been cited in multiple peer-reviewed platforms and contributes to interdisciplinary applications in computer science and medical technology. Her work has garnered 183 citations, reflecting her contributions to the academic community (E-mail: rp.22703019@student.hindustanuniv.ac.in).

Dr. Anitha S Pillai is a Professor at the School of Computing Sciences, Hindustan Institute of Technology and Science (HITS), India and has over 29 years of teaching and research experience. She did her post graduate studies from NIT Calicut and specializes in Artificial Intelligence, Machine Learning, Healthcare Analytics, and Natural Language Processing. She has authored and co-authored over 93 publications, including journal articles, conference papers, and book chapters. Her research contributions have garnered over 943 citations, reflecting her significant impact in the field. In addition to her publications, Dr. Pillai has also co-edited books. Her research significantly advances the application of AI and machine learning in healthcare, with a focus on neuroimaging and data analytics (E-mail: anithasp@hindustanuniv.ac.in).

Dr. Lazzaro di Biase is a Clinical Neurologist with a PhD in the Science of Aging and Tissue Regeneration from Campus Bio-Medico University (UCBM) of Rome. He is the Scientific Director of the Brain Innovations Lab, a UCBM spin-off focused on cutting-edge research in neurodegenerative diseases and neuromodulation technologies. He has gained international clinical and research experience at the University of Oxford, University College London, and Toronto Western Hospital. Dr. di Biase has contributed to advancements in tremor classification, Parkinson's disease monitoring, and adaptive closed-loop therapies. He is the author of more than a dozen peer-reviewed publications, with work appearing in Frontiers in Neurology, Sensors, Expert Review of Neurotherapeutics, and IEEE Transactions on Neural Systems and Rehabilitation Engineering. He also holds four patents and is the founder of a neurotech startup (E-mail: l.dibiase@policlinicocampus.it).

Dr. Ajith Abraham is a globally renowned researcher in artificial intelligence, with over 35 years of multidisciplinary experience. He has authored or co-authored more than 1,500 research publications, including journal articles, conference papers, and books, with some works translated into Chinese, Russian, and Japanese. He has delivered over 250 conference plenary talks and tutorials in more than 20 countries. His scholarly impact is reflected by 63,000+ citations and an H-index of over 119 on Google Scholar. Approximately 1,400 of his papers are indexed in Scopus and over 1,000 in the Web of Science. He has delivered more than 250 plenary talks and tutorials in 20+ countries and served as Chief Editor of Engineering Applications of Artificial Intelligence (EAAI) (Elsevier) from 2016–2021. He sits on the editorial

boards of over 15 international journals and is continuously listed among the top 2% of most cited scientists globally by Stanford/Elsevier. As of 2024, ScholarGPS ranks him among the top 0.01% most cited scientists in engineering and computer science. Currently he is the Vice Chancellor at Sai University, Chennai and the Founding Dean of the School of Artificial Intelligence. Prior to this appointment he was the Vice Chancellor at Bennett University (Times Group) and Dean of Faculty of Computing and Data Science at FLAME University, Pune India. He was the Founding Director of Machine Intelligence Research Labs (MIR Labs—http://www.mirlabs.org), which has members from 100+ countries. He was the Chair of IEEE Systems Man and Cybernetics Society Technical Committee on Soft Computing (2008–2021) and a Distinguished Lecturer of IEEE Computer Society representing Europe (2011–2013). He serves on the editorial board of over 50 International journals (E-mail: ajith.abraham@ieee.org).