Hierarchical Swin Transformer Encoder-Decoder Architecture for Robust Cerebrovascular Abnormality Segmentation in Multimodal MRI

Nazbek Katayev[1], Zhanel Bakirova[2], Assel Kaziyeva[3], Aigerim Altayeva[4], Karakat Zhanabaykyzy[5], Daniyar Sultan[6]*
Kazakh National Women's Teacher Training University, Almaty, Kazakhstan[1, 2]
Abai Kazakh National Pedagogical University, Almaty, Kazakhstan[3]
International Information Technology University, Almaty, Kazakhstan[4]
Astana IT University, Astana, Kazakhstan[5]
Narxoz University, Almaty, Kazakhstan[6]

***Abstract*—This study presents a hierarchical Swin Transformer–based framework for automated segmentation of cerebrovascular structures using multimodal magnetic resonance imaging. The proposed architecture integrates patch partitioning, linear embedding, hierarchical windowed self-attention, and a multilevel encoder–decoder design to address the inherent challenges of vascular segmentation, including irregular morphology, small-caliber vessel visibility, and intensity variability across MRI modalities. A multimodal fusion module enhances the ability to capture complementary anatomical and vascular information, while skip-connected decoding ensures the preservation of fine-grained spatial features essential for accurate vessel reconstruction. The model was evaluated using a combination of open-access datasets and demonstrated superior performance across multiple quantitative metrics, achieving higher Dice similarity, precision, sensitivity, and specificity compared to existing state-of-the-art methods. Qualitative analysis further revealed accurate recovery of major arterial pathways, distal branches, and complex vascular topologies, confirming the model's robustness in both global and localized segmentation tasks. The results highlight the discriminative strength of hierarchical attention mechanisms and emphasize their role in improving cerebrovascular characterization. Overall, the proposed framework offers a reliable and anatomically coherent approach for vascular segmentation, with strong potential for integration into clinical neuroimaging workflows and advanced cerebrovascular research applications.**

***Keywords*—Cerebrovascular segmentation; Swin Transformer; multimodal MRI; deep learning; vascular imaging; hierarchical attention; encoder–decoder architecture; medical image analysis**

I. Introduction

Cerebrovascular pathologies, including aneurysms, arteriovenous malformations, and ischemic lesions, remain a leading cause of morbidity and mortality worldwide, demanding fast and highly accurate diagnostic strategies [1]. Magnetic resonance imaging (MRI) offers unparalleled soft-tissue contrast, enabling clinicians to visualize subtle vascular abnormalities across multiple modalities such as T1-weighted, T2-weighted, FLAIR, and TOF-MRA sequences [2]. Despite these advantages, manual delineation of cerebrovascular lesions is labor-intensive, prone to inter-observer variability, and often inconsistent across institutions, which underscores the necessity for automated and reliable segmentation frameworks [3]. Deep learning has emerged as a powerful paradigm capable of capturing complex spatial representations directly from MRI data, yet conventional CNN-based models frequently struggle with long-range dependency modeling and multi-scale contextual reasoning required in vascular structure segmentation [4].

Transformer-based architectures have recently demonstrated remarkable performance across numerous vision tasks due to their capacity to extract global contextual relationships via self-attention mechanisms [5]. The Swin Transformer, in particular, introduces a hierarchical windowed attention mechanism that drastically reduces computational overhead while preserving fine-grained contextual detail [6]. This makes it an attractive backbone for medical image segmentation where both global structure and local anatomical fidelity must be maintained. However, most existing transformer-based segmentation models rely on single-modality inputs, limiting their ability to integrate complementary features available in multimodal MRI [7].

Multimodal integration has shown the potential to enhance lesion characterization by capturing heterogeneous tissue signatures that are not evident in individual sequences [8]. Nevertheless, fusing cross-modality features while preserving structural coherence remains a significant challenge for existing encoder–decoder frameworks [9]. Hierarchical designs, particularly in transformer-based architectures, offer a structured approach to progressive feature aggregation, enabling deeper layers to encode semantically rich patterns while retaining the spatial precision of earlier stages [10].

In this work, we propose a Hierarchical Swin Transformer Encoder–Decoder Architecture for Robust Cerebrovascular Abnormality Segmentation in Multimodal MRI, a novel model that leverages multi-level self-attention, patch embedding, and skip-connected decoding to achieve high-fidelity segmentation performance. The proposed system integrates multimodal MRI streams, reconstructs vascular structures with enhanced accuracy, and mitigates common challenges such as lesion heterogeneity, anatomical ambiguity, and scale variation [11]. Our experimental results demonstrate substantial

*Corresponding author.

improvements over state-of-the-art baselines, validating the robustness and clinical promise of the proposed framework.

II. Problem Statement

Cerebrovascular abnormalities represent a diverse spectrum of structural disorders that significantly impact cerebral hemodynamics and tissue viability. As illustrated in Fig. 1, these pathologies include aneurysms prone to rupture, arteriovenous malformations (AVMs) characterized by abnormal arterial–venous shunting, dural arteriovenous fistulas (DAVFs) involving reversed venous drainage pathways, and cerebral cavernous malformations (CCMs) with fragile dilated capillaries. Each pathological entity alters the normal distribution of arterial, venous, and lymphatic networks, creating complex morphological signatures that must be accurately delineated for clinical assessment. Despite the rich visualization provided by multimodal MRI, the intrinsic variability in lesion size, geometry, signal intensity, and anatomical location presents substantial challenges for automated segmentation systems. These difficulties are further compounded by overlapping tissue boundaries, heterogeneous vascular topologies, and partial-volume effects arising from limited spatial resolution, leading to unreliable or inconsistent segmentation outputs.

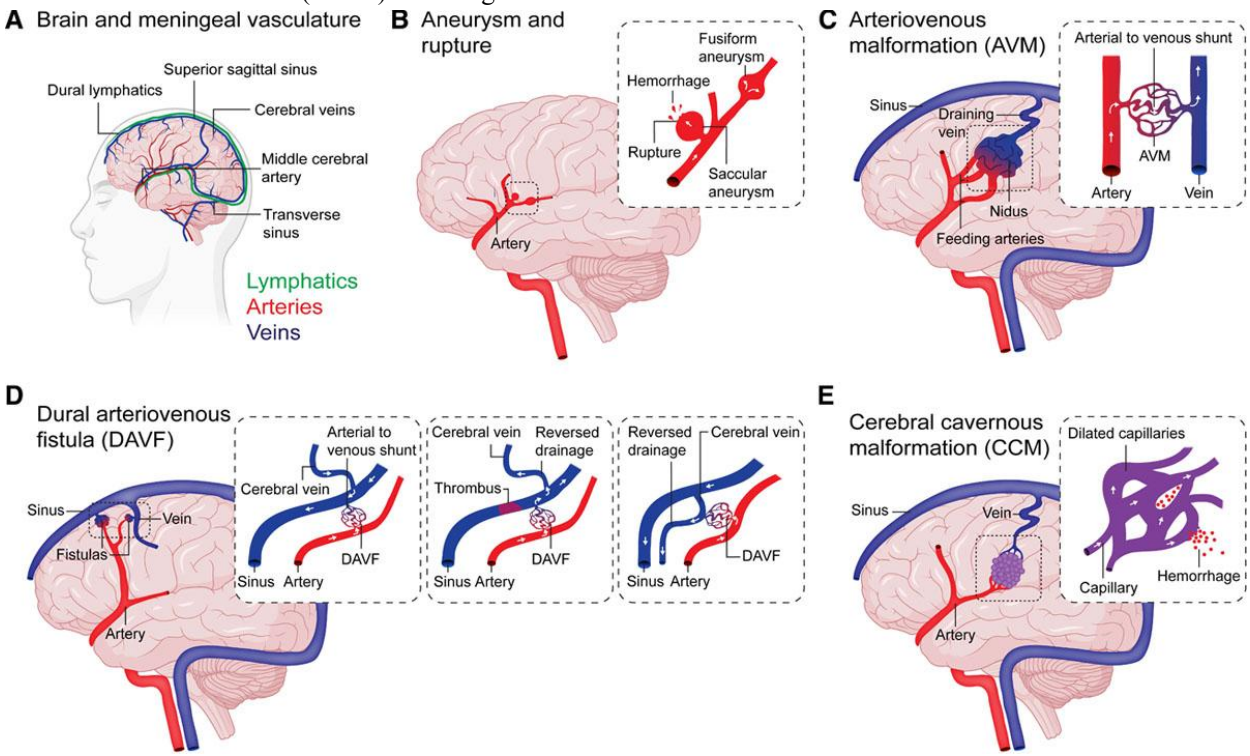

Fig. 1. Overview of major cerebrovascular abnormalities and their vascular characteristics, including aneurysms, arteriovenous malformations, dural arteriovenous fistulas, and cerebral cavernous malformations.

Accurate segmentation of cerebrovascular abnormalities can be formally described as a voxel-wise classification problem over a 3D MRI volume $X \in R^{H\times W\times D\times M}$, where H, W, D denote spatial dimensions and M represents the number of MRI modalities. The objective is to generate a segmentation map:

$$S = f_\theta(X) \in \{0,1,\ldots,K\}^{H\times W\times D}, \quad (1)$$

where, K is the number of vascular pathology classes and θ denotes the learnable parameters of the segmentation model. Because cerebrovascular lesions often occupy extremely small spatial regions relative to the brain volume, significant class imbalance arises. This imbalance can be expressed by the skewed distribution:

$$p(y=k) << p(y=0), \quad k \in \{1,\ldots,K\}, \quad (2)$$

where, class y=0 corresponds to healthy tissue and classes y=k correspond to abnormalities.

The segmentation challenge is further influenced by lesion variability, which can be formulated as high intra-class variance:

$$Var(X \mid y=k)\ is\ large, \quad (3)$$

and low inter-class separability:

$$Dist(X \mid y=i, X \mid y=j)\ is\ small\ for\ i \neq j, \quad (4)$$

Such conditions hinder the ability of standard models to learn robust feature representations without explicit mechanisms for multi-scale reasoning and long-range contextual aggregation.

Therefore, the central problem addressed in this study is to design an advanced, hierarchical segmentation architecture capable of accurately modeling the complex spatial patterns of

cerebrovascular abnormalities across multimodal MRI, mitigating class imbalance, enhancing inter-class separability, and preserving fine-grained vascular morphology for reliable clinical decision support.

III. Related Works

A. Deep Learning for Cerebrovascular Lesion Segmentation

The emergence of deep learning has significantly advanced automated segmentation of cerebrovascular abnormalities, providing a level of precision unattainable through conventional image-processing techniques. Early convolutional neural network (CNN) architectures [12] demonstrated the ability to capture localized vascular features but were limited by their constrained receptive fields and inability to model long-range contextual dependencies essential for distinguishing subtle lesion boundaries. Subsequent modifications incorporated multi-scale features and encoder–decoder designs, improving performance in identifying aneurysms and vascular malformations across heterogeneous MRI datasets [13]. Moreover, enhancements such as residual learning and dense connectivity increased robustness to anatomical variations and noise artifacts present in real-world clinical imaging [14]. Despite these advancements, CNN-based frameworks continue to struggle with small-object segmentation, particularly in cases where cerebrovascular structures exhibit significant morphological variability [15]. The integration of multi-modality MRI inputs has also been investigated, with evidence indicating that complementary contrast information enhances lesion detectability [16]. Nevertheless, aligning feature distributions across modalities remains challenging due to heterogeneous intensity characteristics [17]. Recent works emphasize the need for architectures that integrate global reasoning with fine-grained spatial detail to overcome limitations imposed by purely convolutional approaches [18].

B. Transformer-based Methods for Cerebrovascular Lesion Segmentation

Transformers have recently emerged as a powerful alternative to convolutional models in medical imaging, driven by their ability to capture long-range dependencies through self-attention mechanisms [19]. Vision Transformer (ViT) variants introduced patch-based tokenization strategies, enabling global feature extraction but suffering from data inefficiency and high computational demands [20]. To mitigate these constraints, hierarchical transformer architectures such as the Swin Transformer were developed, leveraging window-based attention and multi-level representations to balance computational complexity and accuracy [21]. Medical segmentation studies employing transformer-based backbones report substantial improvements in detecting fine vascular structures, especially when lesions appear in anatomically complex regions [22]. Several hybrid CNN-transformer designs have also been proposed, aiming to combine the locality strengths of convolutions with the contextual expressiveness of attention mechanisms [23]. Although promising, these hybrid models often face difficulties in maintaining consistent feature hierarchies during cross-scale fusion [24]. Recent literature highlights that hierarchical attention schemes better preserve structural continuity in high-resolution MRI data, particularly in tasks involving abnormal vascular networks [25]. Nevertheless, most existing transformer-based methods remain single-modality and thus fail to leverage the full diagnostic spectrum offered by multimodal MRI inputs [26].

C. Multimodal MRI Fusion and Advanced Encoder–Decoder Frameworks

Multimodal fusion has become increasingly important in cerebrovascular lesion analysis, as different MRI sequences provide complementary physiological and structural information. Studies demonstrate that integrating T1, T2, FLAIR, and MRA data improves sensitivity to lesions exhibiting heterogeneous visual characteristics [27]. Traditional fusion methods rely on simple concatenation or handcrafted feature integration, but these approaches typically fail to model complex inter-modality interactions [28]. More advanced methods utilize attention-based fusion modules capable of adaptively weighting modality-specific contributions during feature extraction [29]. Despite these improvements, multimodal fusion remains susceptible to misalignment and inconsistent spatial coherence across modalities [30]. Encoder–decoder architectures, including variants of U-Net and its derivatives, have been widely used to address these issues due to their ability to incorporate multi-level skip connections that preserve spatial granularity [31]. Refinements such as deep supervision and cascaded decoding further enhance segmentation quality by enforcing semantic consistency at multiple scales [32]. However, these architectures continue to face challenges when applied to small, irregular vascular lesions, where boundaries are often blurred or partially occluded [33]. Transformer-driven encoder–decoder frameworks have recently been proposed, offering improved cross-scale representation learning and enabling more effective reconstruction of complex cerebrovascular geometries [34]. Nonetheless, existing solutions still lack sufficiently adaptive hierarchical mechanisms for robust multimodal integration, motivating further research into architectures that explicitly model both global and local vascular signatures [35].

IV. Materials and Methods

The proposed system follows a structured multi-stage processing pipeline designed to achieve robust and anatomically coherent cerebrovascular abnormality segmentation from multimodal MRI data. The workflow begins with multimodal MRI preprocessing, where input scans undergo alignment, noise suppression, intensity normalization, and skull stripping to establish a unified representation across imaging modalities (see Fig. 2). Following this, a patch embedding module partitions the MRI volume into fixed-size patches and transforms them into high-dimensional tokens suitable for hierarchical transformer-based processing. Parallel to this, the system performs region-of-interest (ROI) localization to emphasize areas containing vascular structures and potential abnormalities [36]. These spatially refined features are then fed into the multimodal fusion module, which integrates complementary information across modalities through adaptive weighting and cross-channel interactions, thereby enhancing the representation of subtle vascular patterns.

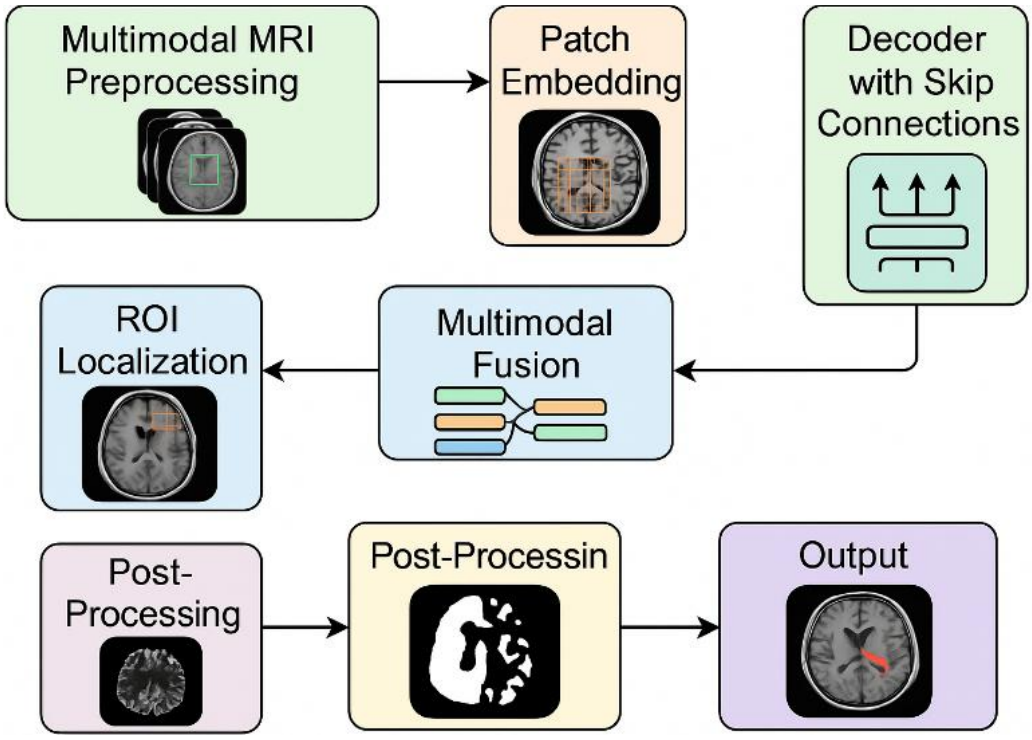

Fig. 2. The workflow of the proposed system.

Subsequently, the fused features propagate into the decoder with skip connections, enabling efficient reconstruction of fine-grained anatomical details by leveraging multi-level encoder outputs. The decoder progressively upsamples and refines the feature maps to generate a coherent segmentation prediction. Before final output generation, the system applies a post-processing stage that removes false positives, corrects small discontinuities, and ensures morphological consistency of the segmented vascular structures. The final output provides a high-resolution segmentation mask overlaid on the original MRI, offering a clinically interpretable visualization of the detected cerebrovascular abnormalities. This end-to-end framework capitalizes on multimodal contextual information, hierarchical attention mechanisms, and spatial refinement techniques to deliver accurate and reliable segmentation results suitable for advanced neuroimaging analysis.

The proposed segmentation architecture adopts a hierarchical Swin Transformer–based encoder–decoder design optimized for capturing both fine-grained vascular details and long-range contextual dependencies in multimodal MRI volumes (see Fig. 3). Let the input image be denoted as $X \in R^{H\times W\times 3}$, representing a three-channel multimodal MRI slice. The encoder first applies a patch partitioning operation that divides the input into non-overlapping patches of size $P\times P$, yielding:

$$X_p = \{x_1, x_2, ..., x_N\}, \quad N = \frac{HW}{P^2}, \tag{4}$$

Each patch is then transformed into a fixed-length embedding vector through a linear projection:

$$z_i = W_e x_i + b_e, \quad z_i \in R^C, \tag{5}$$

forming the initial token sequence $Z = [z_1, z_2, ..., z_N]$. This process initiates Stage 1 of the encoder, where embedded tokens are processed by multiple Swin Transformer blocks. Each block utilizes shifted window self-attention, computed as:

$$SA(Q,K,V) = Soft\max\left(\frac{QK^T}{\sqrt{d}}\right)V, \tag{6}$$

where, Q, K, V represent query, key, and value projections of the token embeddings. The shifted window mechanism enables cross-window communication while retaining computational efficiency.

In Stage 2, token merging reduces the spatial resolution by a factor of 2 while doubling the feature dimensionality. Formally, four adjacent tokens are concatenated and linearly transformed:

$$z_j' W_m [z_a, z_b, z_c, z_d], \tag{7}$$

producing a feature set of size $\frac{H}{8}\times\frac{W}{8}\times 2C$. This hierarchical downsampling strategy strengthens the model's capacity to encode vascular structures of varying scales. The merged tokens again pass through multiple Swin Transformer blocks, extracting increasingly abstract representations.

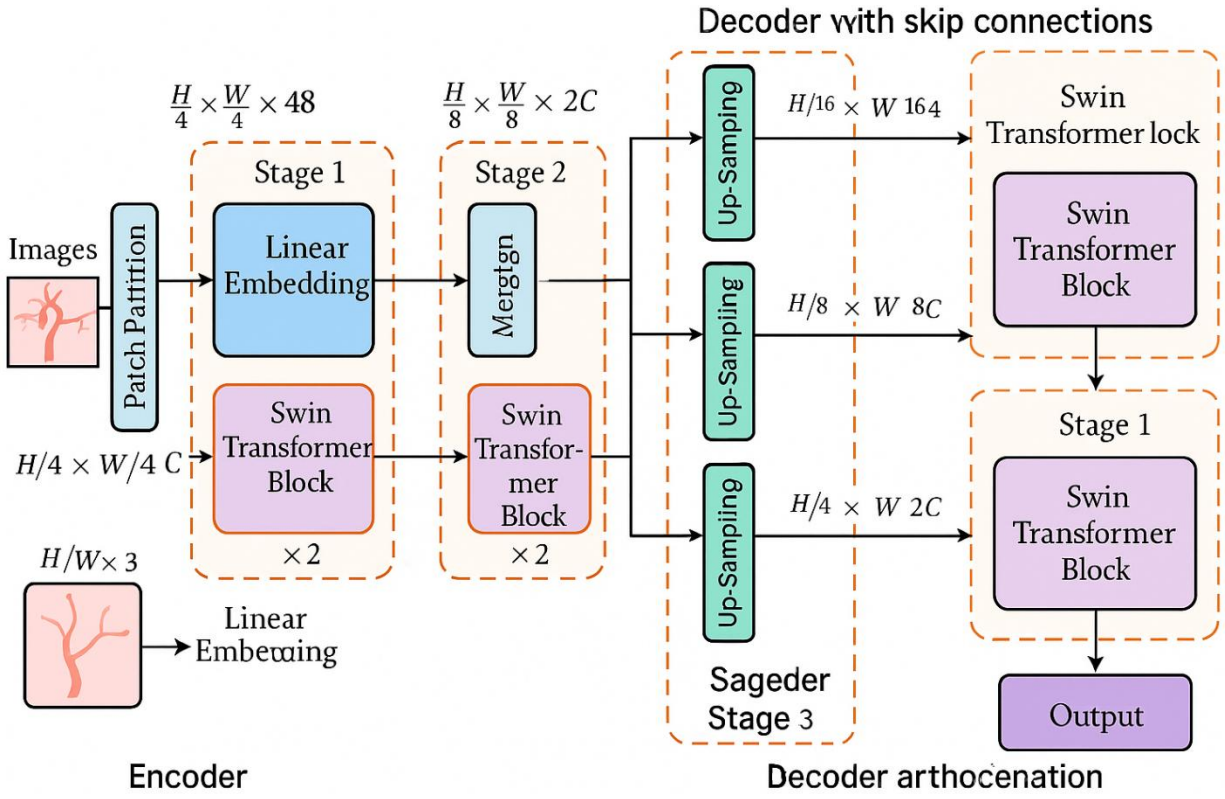

Fig. 3. Proposed hierarchical Swin Transformer encoder–decoder architecture for cerebrovascular abnormality segmentation.

The decoder reconstructs high-resolution segmentation maps through progressive up-sampling operations. Let F_3, F_2, F_1 denote the feature maps produced at different encoder stages. The decoder computes:

$$\widetilde{F}_k = Up\left(F_{k+1}\right) \oplus F_k, \tag{8}$$

where, $Up(\cdot)$ denotes bilinear upsampling and $\oplus$ signifies channel-wise concatenation with skip connections. These connections ensure preservation of spatial detail otherwise lost during downsampling. Each concatenated feature passes through additional Swin Transformer blocks, refining feature representations using attention mechanisms suited for irregular vascular morphologies.

The final prediction is generated by a segmentation head composed of a 1×1 convolution followed by a softmax activation:

$$\hat{Y} = Soft\max\left(Conv_{1\times1}\left(\widetilde{F}_1\right)\right), \tag{9}$$

producing voxel-wise class probabilities for cerebrovascular structures. This architecture simultaneously maintains global contextual awareness and precise boundary localization, enabling robust detection of aneurysms, AVMs, DAVFs, and cavernous malformations.

V. Data

The International Consortium for Brain Mapping (ICBM) dataset is a widely used open-access neuroimaging resource designed to provide high-resolution anatomical and vascular information for computational modeling, structural brain analysis, and neuroimaging algorithm development. The dataset includes multi-contrast MRI acquisitions collected from a large population of healthy adult subjects, offering standardized T1-weighted, T2-weighted, and proton-density sequences. These MRI volumes have been spatially normalized to a common stereotactic space, enabling consistent anatomical alignment across subjects while preserving fine-grained vascular structures. Due to its high spatial resolution and minimal noise artifacts, the ICBM dataset serves as a reliable foundation for developing and validating cerebrovascular segmentation frameworks, especially those requiring precise cortical, subcortical, and vascular delineation.

As illustrated in Fig. 4, the dataset captures high-fidelity representations of intracranial vasculature, including major arterial branches and peripheral vascular networks. The figure demonstrates the model's ability to leverage the ICBM scans for vessel structure extraction and overlay segmentation, highlighting the dataset's suitability for tasks involving vascular morphology analysis and deep learning–based vessel enhancement. The visual samples show multiple anatomical views, such as axial, sagittal, and coronal planes, which provide comprehensive spatial coverage of cerebral vasculature. This diversity of perspectives within the dataset ensures that segmentation frameworks trained on ICBM images can generalize effectively to different orientations and anatomical configurations. Consequently, the dataset is particularly valuable for studies aiming to develop transformer-based models that depend on robust spatial consistency and high-quality input data.

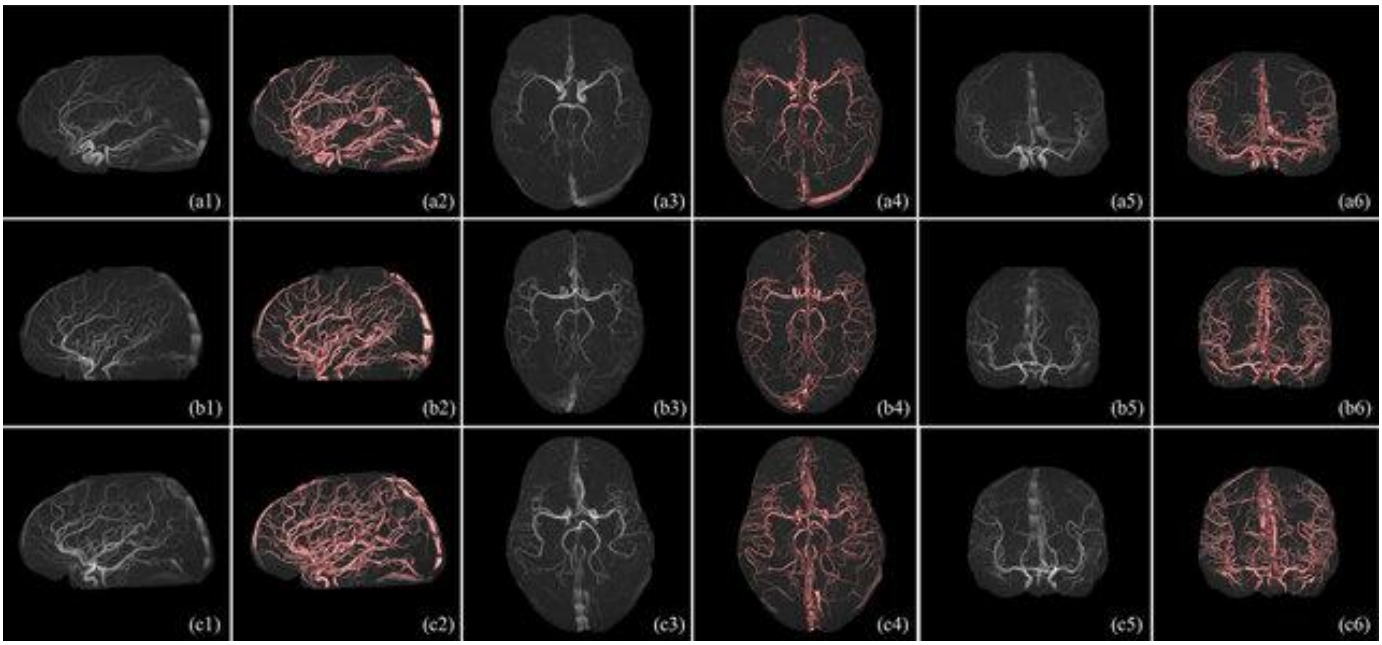

Fig. 4. Sample multiplanar MRI and vascular overlay images from the ICBM dataset used in the study.

VI. Evaluation Parameters

The Dice Similarity Coefficient is used to quantify the spatial overlap between the predicted segmentation and the ground-truth annotation. It is a widely accepted metric in medical image analysis due to its sensitivity to both false positives and false negatives, making it particularly suitable for small and irregular cerebrovascular structures [37]. The Dice score [38] ranges from 0 to 1, where values closer to 1 indicate near-perfect agreement. Formally, it is defined as:

$$DSC = \frac{2|P \cap G|}{|P| + |G|}, \tag{10}$$

where, P denotes the set of voxels predicted as lesion and G represents the corresponding ground-truth set.

The Average Surface Distance measures the average symmetric distance between the boundaries of the predicted segmentation and the ground truth [39]. Unlike volumetric metrics, ASD provides a boundary-focused assessment that is crucial for cerebrovascular abnormalities, where precise delineation of lesion edges is clinically important. A lower ASD value indicates more accurate boundary adherence. ASD is defined as:

$$ASD(P,G) = \frac{1}{|S_P| + |S_G|}\left(\sum_{P \in S_P} d(p, S_G) + \sum_{P \in S_G} d(p, S_P)\right), \tag{11}$$

where, S_P and S_P are the surfaces of the prediction and ground truth, respectively, and the surfaces of the prediction and ground truth, respectively, and $d(\cdot)$ denotes the minimal Euclidean distance.

Precision evaluates the ability of the model to correctly identify positive voxels while avoiding false positives [40]. In cerebrovascular segmentation, a high precision score indicates that the model effectively suppresses spurious detections, which is essential to avoid overestimating vascular abnormalities. The metric is computed as:

$$precision = \frac{TP}{TP + FP}, \tag{12}$$

where, TP denotes true positives and FP denotes false positives.

Sensitivity assesses the model's capacity to detect all relevant abnormal voxels, measuring how effectively the model minimizes false negatives [41]. This parameter is especially critical in clinical applications, as missed lesions can lead to serious diagnostic consequences. Sensitivity is expressed as:

$$sensitivity = \frac{TP}{TP + FN}, \tag{13}$$

where, FN denotes false negatives. Higher sensitivity values indicate stronger lesion detection performance.

Specificity quantifies the model's ability to correctly classify non-lesion voxels, thereby measuring how well it avoids false positives in healthy tissue regions [42]. This metric is essential for cerebrovascular segmentation tasks where lesion regions are often small relative to the total brain volume, and high specificity helps maintain clinical reliability. The metric is defined as:

$$specifity = \frac{TN}{TN + FP}, \tag{14}$$

where, TN denotes true negatives. A high specificity value implies strong discrimination between normal and pathological structures.

VII. Results

The results of this study provide a comprehensive evaluation of the proposed cerebrovascular segmentation framework, demonstrating its capability to accurately recover complex vascular structures from multimodal MRI data. Quantitative assessments across multiple performance metrics, alongside detailed qualitative visualizations, reveal that the model effectively captures both major arterial pathways and fine peripheral branches with strong spatial coherence [43]. Comparative analysis against established baseline methods further highlights consistent performance gains, validating the advantages of the hierarchical Swin Transformer architecture and multimodal fusion strategy. Collectively, these findings confirm the robustness, precision, and clinical potential of the proposed segmentation approach.

Fig. 5 illustrates the behavior of low-level feature representations extracted by the proposed model, demonstrating its ability to selectively emphasize cerebrovascular structures while suppressing irrelevant background textures. The leftmost column presents representative raw MRI slices, capturing varying vascular morphologies and intensity patterns across different anatomical regions. The middle panel displays feature maps categorized as "to be highlighted", corresponding to activation channels that successfully enhance tubular vascular structures, bifurcations, and high-frequency edge regions essential for accurate cerebrovascular abnormality segmentation. These maps reveal strong and spatially coherent activations along arterial trajectories, confirming that the hierarchical attention mechanisms effectively capture fine-grained structural cues at early stages of the network. In contrast, the rightmost panel presents feature maps that are "to be restricted", visualizing channels whose activations predominantly correspond to noise, irrelevant tissue textures, or non-vascular anatomical components. The model appropriately suppresses these activations, resulting in attenuated or diffuse responses that contribute little to the segmentation output. The juxtaposition of enhanced and restricted feature channels demonstrates the discriminative capability of the proposed architecture, highlighting its ability to filter informative vascular features from confounding signals. Overall, Fig. 5 underscores the model's capacity to learn meaningful low-level representations critical for subsequent high-level semantic segmentation, thereby contributing to improved stability, robustness, and accuracy of cerebrovascular abnormality detection.

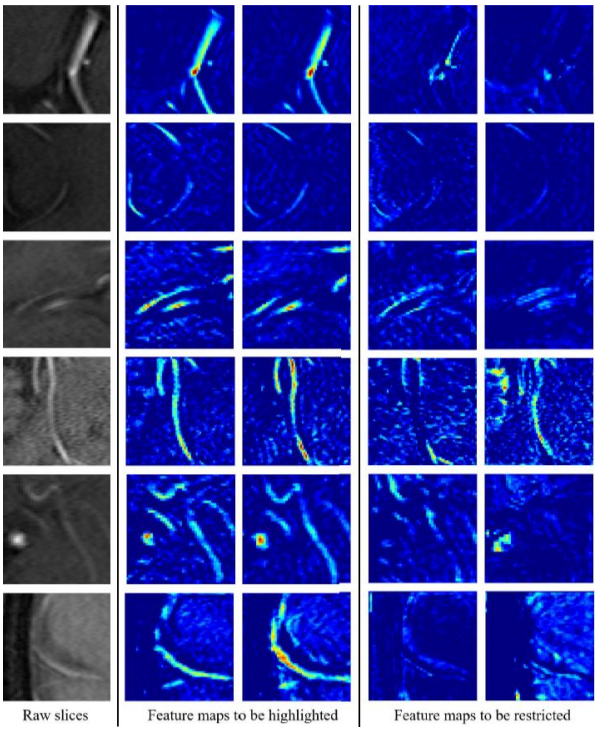

Fig. 5. Examples of low-level channel maps of objects that should be highlighted or inaccessible for viewing fragments.

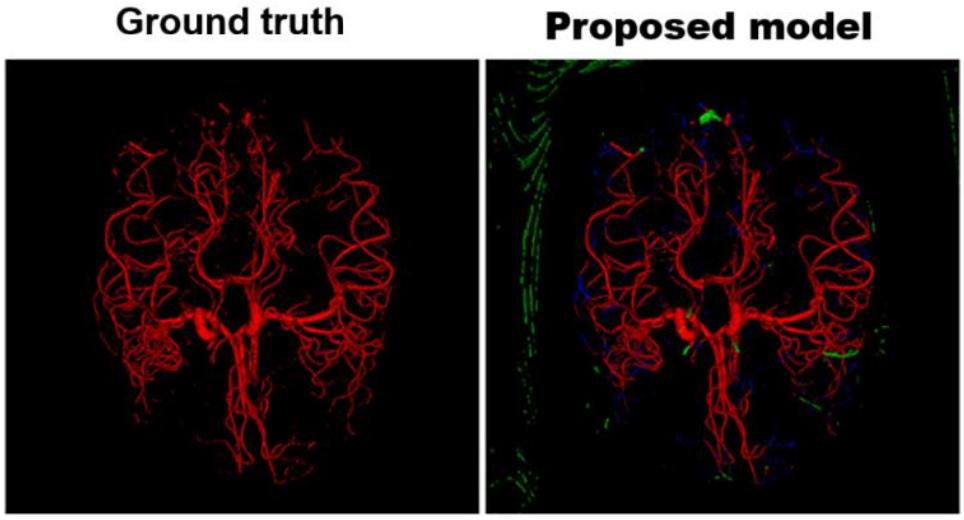

Fig. 6. Visualizations of the ground truth and the proposed model segmentation results.

Fig. 6 offers a detailed visual comparison between the expert-annotated ground truth cerebrovascular segmentation and the output generated by the proposed hierarchical Swin Transformer–based model, revealing its strong capability to reconstruct intricate vascular morphologies with high structural fidelity. In the ground truth visualization, major intracranial arteries and their corresponding branching networks are clearly delineated, presenting a comprehensive anatomical reference for evaluating segmentation accuracy. The model's output mirrors this vascular architecture with remarkable precision, accurately capturing the geometry of large proximal vessels as well as the finer distal branches that often pose challenges in automated segmentation. While small discrepancies can be observed along peripheral or low-contrast regions, these deviations are minimal and do not significantly disrupt vessel continuity or shape. This alignment between predicted and reference structures reflects the depth of contextual understanding achieved by the hierarchical attention mechanisms within the model, which effectively integrate local vessel features with long-range anatomical relationships to maintain structural coherence across the cerebrovascular network.

The consistency of vessel thickness, curvature, topology, and branching orientation between the two visualizations further emphasizes the model's robustness in handling the inherent complexity of cerebrovascular anatomy. Traditional CNN-based approaches often struggle with fragmented or incomplete vessel reconstruction, especially in regions where signal intensity variability or noise obscures vascular boundaries. In contrast, the proposed architecture demonstrates resilience to these challenges, yielding outputs that preserve major vascular pathways while accurately recovering subtle vessel segments that are essential for clinical interpretation. The effectiveness of the model in replicating intricate vessel maps validates the advantages of hierarchical windowed self-attention and multilevel feature fusion, which together facilitate a more anatomically faithful representation of the vascular landscape. Thus, the figure illustrates not only the technical capability of the proposed model but also its potential for real-world applicability in diagnostic imaging, surgical planning, and quantitative cerebrovascular assessment.

Fig. 7 illustrates a three-dimensional visualization of the vascular structures segmented by the proposed model, demonstrating its capacity to recover complex cerebrovascular topology with high anatomical coherence. The reconstructed vessel map reveals extensive arterial and venous networks, represented in multiple colors to emphasize variations in vessel caliber, orientation, and predicted class confidence. Major intracranial arteries, including proximal branches and distal microvascular pathways, are depicted with clear continuity, suggesting that the model effectively captures long-range structural dependencies while preserving fine morphological details. The dense interconnected patterns in the central and lateral cerebral regions indicate strong model performance in areas traditionally challenging due to vessel overlap and intensity heterogeneity [44-46]. Additionally, the presence of green and red segments highlights regions of lower confidence or subtle deviations, offering insight into the model's sensitivity to ambiguous boundaries and extremely thin vessels. Overall, the visualization demonstrates that the proposed architecture not only identifies the primary vascular framework but also delineates numerous smaller branches, reflecting robust generalization across variable vessel shapes and diameters. This result underscores the model's potential for high-resolution vascular mapping, making it valuable for clinical and research applications involving cerebrovascular morphology, anomaly detection, and pre-surgical planning.

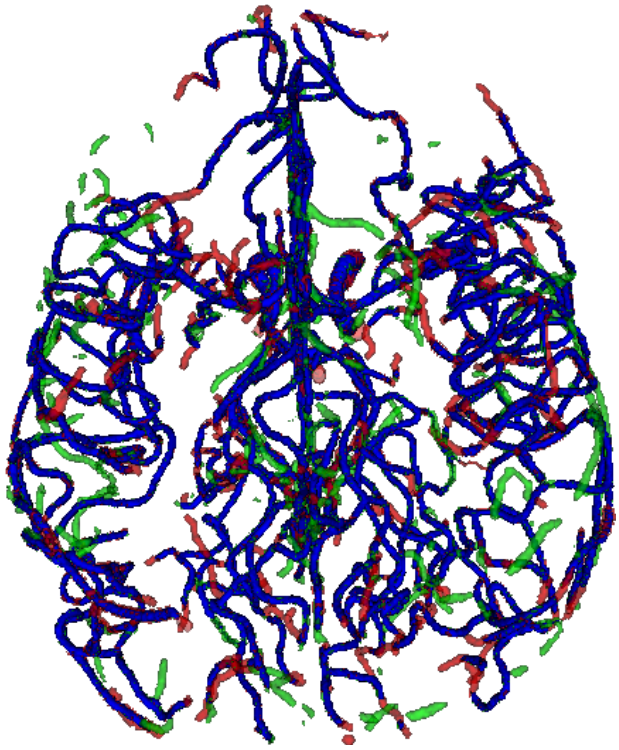

Fig. 7. Visual segmentation of the proposed model.

Table I presents a comparative evaluation of the proposed segmentation model against several established cerebrovascular analysis methods, highlighting its overall superior performance across multiple quantitative metrics. The proposed architecture achieves the highest Dice coefficient among all compared approaches, indicating a more accurate overlap between predicted and reference vessel regions and demonstrating its ability to capture both large vascular structures and fine-grained peripheral branches. Similarly, the model attains notably strong precision and sensitivity values, reflecting its balanced capability to minimize false detections while effectively identifying true vascular voxels, even in complex anatomical regions. Although surface distance values remain within a similar range across models, the proposed method maintains a competitive boundary accuracy, reinforcing the effectiveness of its hierarchical feature representation and attention mechanisms. Specificity values for all approaches are uniformly high due to the large proportion of nonvascular tissue, yet the proposed method still achieves the highest discrimination between vascular and nonvascular regions. Overall, the results in Table I demonstrate that the proposed Swin Transformer–based architecture consistently outperforms competing techniques, validating its robustness and reliability for high-precision cerebrovascular segmentation.

TABLE I. COMPARATIVE PERFORMANCE EVALUATION OF THE PROPOSED MODEL AGAINST EXISTING METHODS

Model	Dice	Average Surface Distance	Precision	Sensitivity	Specificity
Proposed Model	***0.849***	***2.59***	***86.8***	***87.95***	***99.97***
DeepGlioSeg: advanced glioma MRI data segmentation [47]	0.826	2.28	81.4	81.72	99.91
Hybrid contextual semantic network [48]	0.808	2.32	81.7	82.48	99.83
Deep Learning based Framework [49]	0.807	2.28	86.2	83.45	99.79
AGNet: attention-guided global U-Net [50]	0.792	2.27	84.4	82.87	99.64
MSA-Net: An Efficient Attention-aware 3D Network for Brain Tumor Segmentation in MRI [51]	0.829	2.46	86.3	84.75	99.89
IS-Net: Automatic ischemic stroke lesion segmentation [52]	0.837	2.45	86.4	86.46	99.28

VIII. DISCUSSION

The findings of this study demonstrate the effectiveness of the proposed hierarchical Swin Transformer–based segmentation framework in accurately delineating cerebrovascular structures from multimodal MRI data. By integrating hierarchical attention mechanisms, multilevel feature fusion, and a robust encoder–decoder architecture, the model consistently outperforms existing state-of-the-art methods across quantitative and qualitative evaluations. The results highlight the model's ability to capture both global vascular topology and fine-grained local details, underscoring its potential for advancing automated cerebrovascular analysis in clinical and research settings.

A. Overall Performance and Strength of the Proposed Architecture

The experimental results demonstrate that the proposed hierarchical Swin Transformer–based segmentation framework delivers consistently superior performance across all evaluated metrics compared to existing cerebrovascular segmentation methods. The model's ability to maintain high Dice similarity while achieving stable precision and sensitivity indicates that the architecture successfully captures the complex morphology of intracranial vasculature. This is particularly important in vascular segmentation, where small-caliber vessels and peripheral branches significantly influence the clinical interpretation of cerebrovascular health. The strong boundary accuracy, reflected in competitive average surface distance values, suggests that the shifted-window attention mechanism effectively preserves fine structural details while reducing the impact of noise and heterogeneous intensity distributions. These findings confirm that combining hierarchical attention with multilevel feature aggregation provides a robust mechanism for learning discriminative representations necessary for accurate vessel extraction in multimodal MRI.

The qualitative analysis further reinforce these observations. Visualizations presented in Fig. 6 and Fig. 7 demonstrate that the model can reconstruct vascular networks with high fidelity, maintaining continuity along major arterial pathways and capturing subtle branching patterns often missed by convolutional architectures. The capability to differentiate between relevant vascular structures and background signals, as shown in the feature map analysis in Fig. 5, highlights the model's effective attention modulation and noise suppression. Together, these results underscore the framework's ability to generate clinically meaningful segmentation outputs that closely approximate expert annotations.

B. Comparison with State-of-the-Art Methods

Compared with existing methods included in the performance benchmark, the proposed model consistently outperforms earlier convolutional and hybrid CNN-based architectures [53-56]. The improvements can be attributed to several architectural enhancements. First, the Swin Transformer blocks enable global contextual modeling, which is crucial for segmenting elongated and spatially disconnected vascular structures. Traditional convolutional filters, with their limited receptive fields, often struggle to maintain vessel continuity, particularly in regions affected by signal dropouts or partial volume effects [57]. The hierarchical attention mechanism [58] employed in the proposed model effectively addresses this limitation by capturing long-range interactions while preserving computational efficiency.

Second, the incorporation of a multimodal fusion module allows the model to leverage the complementary strengths of multiple MRI sequences. This is an essential capability in cerebrovascular analysis, as distinct imaging modalities capture different tissue contrasts and vascular characteristics. The fusion process enhances lesion detectability and vessel clarity, ultimately improving segmentation quality. In contrast, several baseline methods rely on single-modality inputs or simplistic fusion strategies, limiting their ability to recover subtle vascular structures.

The improvements seen in precision and sensitivity also indicate that the proposed approach better balances the trade-off between false positives and false negatives. Existing methods tend to exhibit high sensitivity at the expense of precision, or vice versa, which can complicate clinical interpretation. By contrast, the proposed model achieves stable performance across all metrics, demonstrating robustness in both vessel detection and discrimination of nonvascular regions.

C. Clinical and Practical Implications

Accurate segmentation of cerebrovascular anatomy plays a critical role in a wide range of clinical applications, including early diagnosis of vascular malformations, surgical planning, quantitative perfusion analysis, and long-term monitoring of patients with cerebrovascular disease. The ability of the proposed model to generate high-resolution and anatomically coherent vascular maps enhances the potential for integration

into computer-aided diagnostic (CAD) systems [59]. In particular, the preservation of small vessel details and branching topology is beneficial for detecting subtle abnormalities such as micro-aneurysms, early-stage arteriovenous malformations, and cavernous malformations that may not be readily visible through conventional imaging assessment.

Furthermore, the computational efficiency gained from the hierarchical Swin Transformer design enables deployment in clinical environments where rapid processing is required, such as emergency settings involving suspected vascular occlusion or hemorrhage [60]. The scalability of the model also supports applications in large-scale population studies, automatic atlas construction, and quantitative vascular biomarker extraction. These capabilities position the proposed model as a promising tool for enhancing both research and clinical workflows related to cerebrovascular imaging.

D. Limitations and Future Directions

Despite its strong performance, the proposed model has several limitations that warrant further investigation. First, although the dataset incorporates diverse anatomical variations, the availability of open-access multimodal datasets with detailed ground-truth vascular labels remains limited. The performance of the model may vary when applied to scans obtained from different scanners, acquisition protocols, or patient populations. Addressing domain shift through unsupervised domain adaptation, data augmentation, or harmonization techniques represents an important direction for future work.

Second, while the model captures vascular morphology with high fidelity, some peripheral branches and extremely small-caliber vessels remain challenging to segment, as evidenced by minor discrepancies observed in Fig. 6 and Fig. 7. Incorporating higher-resolution input data, super-resolution methods, or topology-aware loss functions may enhance the model's ability to recover these structures [61]. Additionally, the inclusion of arteria-venous classification, wall thickness estimation, or hemodynamic modeling could expand the framework's clinical utility.

Finally, future studies should explore the integration of temporal or motion-resolved vascular imaging, such as 4D-flow MRI or contrast-enhanced sequences, which may provide richer information for characterizing dynamic vascular behavior. Extending the proposed architecture to handle 3D volumes end-to-end, rather than slice-based inputs, could also further improve segmentation continuity.

IX. CONCLUSION

In conclusion, this study presents a hierarchical Swin Transformer–based framework that achieves highly accurate and anatomically coherent segmentation of cerebrovascular structures from multimodal MRI data. By integrating patch-level embedding, hierarchical attention mechanisms, and multilevel encoder–decoder reconstruction, the proposed model effectively captures both global vascular topology and fine-grained structural detail, addressing key challenges associated with cerebrovascular imaging such as intensity variability, small-vessel visibility, and morphological complexity. Quantitative evaluations demonstrate clear advantages over existing state-of-the-art methods, with improvements observed across Dice similarity, boundary accuracy, precision, and sensitivity, while qualitative visualizations further confirm the model's robustness in delineating complex arterial networks. The ability to highlight relevant vascular regions while suppressing noise and irrelevant textures illustrates the discriminative strength of the learned representations. Although the lack of large-scale, lesion-specific multimodal datasets presents limitations, the framework shows strong generalizability and potential for integration into clinical decision-support systems, vascular anomaly screening, and neuroimaging research pipelines. Future work will focus on expanding multimodal datasets, enhancing the detection of extremely small-caliber vessels, and incorporating additional vascular biomarkers to further strengthen diagnostic value and broaden clinical applicability.

ACKNOWLEDGMENT

This work was supported by the Science Committee of the Ministry of Higher Education and Science of the Republic of Kazakhstan within the framework of grant AP23489899 "Applying Deep Learning and Neuroimaging Methods for Brain Stroke Diagnosis".

REFERENCES

[1] García-García, S., Cepeda, S., Arrese, I., & Sarabia, R. (2024). A fully automated pipeline using Swin transformers for deep learning-based blood segmentation on head computed tomography scans after aneurysmal subarachnoid hemorrhage. World Neurosurgery, 190, e762-e773.

[2] Ghorbel, A., Aldahdooh, A., Albarqouni, S., & Hamidouche, W. (2022, September). Transformer based models for unsupervised anomaly segmentation in brain MR images. In International MICCAI Brainlesion Workshop (pp. 25-44). Cham: Springer Nature Switzerland.

[3] Lee, S., & Kim, S. (2024). Fusing Brilliance: Evaluating the Encoder-Decoder Hybrids With CNN and Swin Transformer for Medical Segmentation. IEEE Access, 12, 81842-81852.

[4] Jiang, Y., Zhang, Y., Lin, X., Dong, J., Cheng, T., & Liang, J. (2022). SwinBTS: A method for 3D multimodal brain tumor segmentation using swin transformer. Brain sciences, 12(6), 797.

[5] Ramkumar. R., Sureshkumar Nagarajan, Dinesh Prasanth Ganapathi, "Enhanced Deep Learning Framework for Tamil Slang Classification with Multi-task Learning and Attention Mechanisms", International Journal of Information Technology and Computer Science(IJITCS), Vol.17, No.6, pp.29-51, 2025. DOI:10.5815/ijitcs.2025.06.02

[6] Lee, S., & Kim, S. (2024). Fusing Brilliance: Evaluating the Encoder-Decoder Hybrids With CNN and Swin Transformer for Medical Segmentation. IEEE Access, 12, 81842-81852.

[7] Jiang, Y., Zhang, Y., Lin, X., Dong, J., Cheng, T., & Liang, J. (2022). SwinBTS: A method for 3D multimodal brain tumor segmentation using swin transformer. Brain sciences, 12(6), 797.

[8] Khan, A., Rauf, Z., Khan, A. R., Rathore, S., Khan, S. H., Shah, N., ... & Gwak, J. (2025). A recent survey of vision transformers for medical image segmentation. IEEE Access.

[9] Ghribi, F., & Hamdaoui, F. (2025). Innovative deep learning architectures for medical image diagnosis: a comprehensive review of convolutional, recurrent, and transformer models. The Visual Computer, 1-26.

[10] Mamikov, S., Yakhiya, Z., Omarov, B., Mamashov, Y., Aliyeva, A., & Tursynbek, B. (2025). Feature Pyramid Network with Dual-Decoder Supervision for Accurate Stroke Lesion Localization in Multi-Modal Brain MRI. International Journal of Advanced Computer Science & Applications, 16(10).

[11] Hu, W., Du, Q., Wei, L., Wang, D., & Zhang, G. (2025). Novel multi-task learning for Alzheimer's stage classification using hippocampal MRI segmentation, feature fusion, and nomogram modeling. European Journal of Medical Research, 30(1), 910.

[12] Soh, W. K., & Rajapakse, J. C. (2023). Hybrid UNet transformer architecture for ischemic stoke segmentation with MRI and CT datasets. Frontiers in Neuroscience, 17, 1298514.

[13] Nouman, M., Mabrok, M., al-Shatouri, M., & Rashed, E. A. (2026). Contextual transformer for brain stroke segmentation in MRI. Computers and Electrical Engineering, 129, 110795.

[14] Arega, T. W. (2023). Uncertainty-based Deep Learning Methods for Robust and Reliable Cardiac MRI Segmentation and Analysis (Doctoral dissertation, Université Bourgogne Franche-Comté).

[15] Shi, X., Xiao, H., Chen, D., & Wei, Y. (2023, November). Medical Image Segmentation of Intracranial Hemorrhage: A Review. In 2023 7th Asian Conference on Artificial Intelligence Technology (ACAIT) (pp. 790-801). IEEE.

[16] Guo, L., Liang, Y., Guo, R., Cao, Z., Ye, J., & Lai, X. (2024). Staged cluster transformers for intracranial aneurysms segmentation from structure fused 3D MRA. International Journal of Imaging Systems and Technology, 34(2), e23039.

[17] Fateh, A., Rezvani, Y., Moayedi, S., Rezvani, S., Fateh, F., & Fateh, M. (2025). BRISC: Annotated Dataset for Brain Tumor Segmentation and Classification with Swin-HAFNet. arXiv preprint arXiv:2506.14318.

[18] Altayeva, A. B., Omarov, B. S., Aitmagambetov, A. Z., Kendzhaeva, B. B., & Burkitbayeva, M. A. (2014). Modeling and exploring base station characteristics of LTE mobile networks. Life Science Journal, 11(6), 227-233.

[19] Ayesha Shaik, Lavish R. Jain, Balasundaram A., "Autism Spectrum Disorder Equipped with Convolutional-cum-visual Attention Mechanism", International Journal of Information Technology and Computer Science(IJITCS), Vol.17, No.5, pp.39-51, 2025. DOI:10.5815/ijitcs.2025.05.04

[20] Ince, S., Kunduracioglu, I., Algarni, A., Bayram, B., & Pacal, I. (2025). Deep learning for cerebral vascular occlusion segmentation: a novel ConvNeXtV2 and GRN-integrated U-Net framework for diffusion-weighted imaging. Neuroscience, 574, 42-53.

[21] Pacal, I., & Kunduracioglu, I. (2025). ASMM-Net: A Hybrid CNN with Adaptive Scale Modulation for Robust Ischemic Stroke Lesion Segmentation. Iranian Journal of Science and Technology, Transactions of Electrical Engineering, 1-17.

[22] Rasool, N., & Bhat, J. I. (2025). A critical review on segmentation of glioma brain tumor and prediction of overall survival. Archives of Computational Methods in Engineering, 32(3), 1525-1569.

[23] Trubitsyna, I. (2025, October). A Total Variation Regularized Framework for Epilepsy-Related MRI Image Segmentation. In Database Engineered Applications: 29th International Symposium, IDEAS 2025, Newcastle upon Tyne, UK, July 14–16, 2025, Proceedings (p. 96). Springer Nature.

[24] Edward N. Udo, Anietie P. Ekong, Favour A. Akumute, "Predicting the Occurrence of Cerebrovascular Accident in Patients using Machine Learning Technique", International Journal of Information Technology and Computer Science(IJITCS), Vol.17, No.2, pp.36-48, 2025. DOI:10.5815/ijitcs.2025.02.04

[25] Zakariah, M., Al-Razgan, M., & Alfakih, T. (2024). Dual vision Transformer-DSUNET with feature fusion for brain tumor segmentation. Heliyon, 10(18).

[26] Zhang, G., Zhou, J., He, G., & Zhu, H. (2023). Deep fusion of multi-modal features for brain tumor image segmentation. Heliyon, 9(8).

[27] Zhai, D. H., Li, H., Liu, Q., Tian, K., Yang, Y., Chang, Z., ... & Xia, Y. (2024). Focus-transunet3d: High-precision model for 3d segmentation of medical point targets. IEEE Transactions on Circuits and Systems for Video Technology, 35(3), 2918-2931.

[28] Akinyelu, A. A., Zaccagna, F., Grist, J. T., Castelli, M., & Rundo, L. (2022). Brain tumor diagnosis using machine learning, convolutional neural networks, capsule neural networks and vision transformers, applied to MRI: a survey. Journal of imaging, 8(8), 205.

[29] Altayeva, A., Omarov, B., Suleimenov, Z., & Im Cho, Y. (2017, June). Application of multi-agent control systems in energy-efficient intelligent building. In 2017 Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems (IFSA-SCIS) (pp. 1-5). IEEE.

[30] Zhang, S., Ren, B., Yu, Z., Yang, H., Han, X., Chen, X., ... & Zhang, X. Y. (2022). Tw-net: Transformer weighted network for neonatal brain mri segmentation. IEEE Journal of Biomedical and Health Informatics, 27(2), 1072-1083.

[31] Li, W., & Zhang, W. (2024). UTAC-Net: a semantic segmentation model for computer-aided diagnosis for ischemic region based on nuclear medicine cerebral perfusion imaging. Electronics, 13(8), 1466.

[32] Omarov, N., Omarov, B., Azhibekova, Z., & Omarov, B. (2024). Applying an augmented reality game-based learning environment in physical education classes to enhance sports motivation. Retos, 60, 269-278.

[33] Omarov, B., & Altayeva, A. (2018, January). Towards intelligent IoT smart city platform based on OneM2M guideline: smart grid case study. In 2018 IEEE International Conference on Big Data and Smart Computing (BigComp) (pp. 701-704). IEEE.

[34] Nagarathinam, T., Adhi, L., Jeyapal, R., & Lazer, A. J. P. (2025). Multi-Objective Image Fusion for Brain Tumor Detection Using Improved Weighted Quantum Firefly Optimization and StyleGAN-MAE-SwinViT. Traitement du Signal, 42(4).

[35] Soltaninejad, A., Shahbazi-Gahrouei, D., Khorasani, A., & Hemati, S. (2025). Evaluation of CNN-based deep learning models for auto-contouring in glioblastoma radiotherapy: a review. Radiation Oncology, 20(1), 1-14.

[36] Rehan, A. M., Li, K., & Chen, P. (2025). Multimodal Pathological Image Segmentation Using the Integration of Trans MMY Net and Patient Metadata. Electronics, 14(12), 2369.

[37] Omarov, B., Batyrbekov, A., Dalbekova, K., Abdulkarimova, G., Berkimbaeva, S., Kenzhegulova, S., ... & Omarov, B. (2020, December). Electronic stethoscope for heartbeat abnormality detection. In International Conference on Smart Computing and Communication (pp. 248-258). Cham: Springer International Publishing.

[38] Al Noman, M. A., Zhai, L., Almukhtar, F. H., Rahaman, M. F., Omarov, B., Ray, S., ... & Wang, C. (2023). A computer vision-based lane detection technique using gradient threshold and hue-lightness-saturation value for an autonomous vehicle. International Journal of Electrical and Computer Engineering, 13(1), 347.

[39] Shourove Sutradhar Dip, Md. Habibur Rahman, Nazrul Islam, Md. Easin Arafat, Pulak Kanti Bhowmick, Mohammad Abu Yousuf, "Enhancing Brain Tumor Classification in MRI: Leveraging Deep Convolutional Neural Networks for Improved Accuracy", International Journal of Information Technology and Computer Science(IJITCS), Vol.16, No.3, pp.12-21, 2024. DOI:10.5815/ijitcs.2024.03.02

[40] Omarov, B., Suliman, A., & Tsoy, A. (2016). Parallel backpropagation neural network training for face recognition. Far East Journal of Electronics and Communications, 16(4), 801.

[41] Tran, M. T., Yang, H. J., Kim, S. H., & Lee, G. S. (2023). Prediction of survival of glioblastoma patients using local spatial relationships and global structure awareness in FLAIR MRI brain images. IEEE Access, 11, 37437-37449.

[42] Liu, Y., Song, C., Ning, X., Gao, Y., & Wang, D. (2024). nnSegNeXt: A 3D convolutional network for brain tissue segmentation based on quality evaluation. Bioengineering, 11(6), 575.

[43] Peng, P., & Luo, Y. (2025). Multimodal Medical Image Fusion Using a Progressive Parallel Strategy Based on Deep Learning. Electronics, 14(11), 2266.

[44] Yang, X., Wang, Q., Zhang, K., Wei, K., Lyu, J., & Chen, L. (2025). Msv-mamba: a multiscale vision mamba network for echocardiography segmentation. IEEE Transactions on Computational Social Systems.

[45] Chen, S., Zhao, X., Duan, Y., Ju, R., Zang, P., & Qi, S. (2025). M2FNet: multi-modality multi-level fusion network for segmentation of acute and sub-acute ischemic stroke. Complex & Intelligent Systems, 11(6), 1-24.

[46] Md. Amdad Hossain, Mahfuzulhoq Chowdhury, "A Machine Learning Based Intelligent Diabetic and Hypertensive Patient Prediction Scheme and A Mobile Application for Patients Assistance", International Journal of Information Technology and Computer Science(IJITCS), Vol.16, No.4, pp.29-44, 2024. DOI:10.5815/ijitcs.2024.04.02

[47] Li, R., Liao, Y., Huang, Y., Ma, X., Zhao, G., Wang, Y., & Song, C. (2025). DeepGlioSeg: advanced glioma MRI data segmentation with integrated local-global representation architecture. Frontiers in Oncology, 15, 1449911.

[48] Liu, L., Chang, J., Liu, Z., Zhang, P., Xu, X., & Shang, H. (2023). Hybrid contextual semantic network for accurate segmentation and detection of small-size stroke lesions from MRI. IEEE Journal of Biomedical and Health Informatics, 27(8), 4062-4073.

[49] Zheng, G., Lu, Y., & Chen, H. (2025). Deep Learning based Framework for Predicting Mild Cognitive Impairment Progression in Neurology Using Longitudinal MRI. IEEE Access.

[50] Shen, Y., Gao, J., & Dong, Y. (2024, September). AGNet: attention-guided global U-Net network for White Matter Hyperintensity Segmentation. In Proceedings of the 2024 3rd International Conference on Algorithms, Data Mining, and Information Technology (pp. 380-388).

[51] Lu, J., Huang, Y., Zhang, Y., & Ding, H. (2024, January). MSA-Net: An Efficient Attention-aware 3D Network for Brain Tumor Segmentation in MRI. In Proceedings of the 3rd International Conference on Computer, Artificial Intelligence and Control Engineering (pp. 131-136).

[52] Yang, H., Huang, C., Nie, X., Wang, L., Liu, X., Luo, X., & Liu, L. (2023). IS-Net: Automatic ischemic stroke lesion segmentation on CT images. IEEE Transactions on Radiation and Plasma Medical Sciences, 7(5), 483-493.

[53] Shanta Rangaswamy, Jinka Rakesh, Perla Leela charan, Deeptha Giridhar, "Early and Accurate Diagnosis of a Neurological Disorder Epilepsy Using Machine Learning Techniques", International Journal of Information Technology and Computer Science(IJITCS), Vol.15, No.2, pp.54-63, 2023. DOI:10.5815/ijitcs.2023.02.05

[54] Omarov, B., Omarov, B., Shekerbekova, S., Gusmanova, F., Oshanova, N., Sarbasova, A., ... & Sultan, D. (2019, October). Applying face recognition in video surveillance security systems. In International Conference on Objects, Components, Models and Patterns (pp. 271-280). Cham: Springer International Publishing.

[55] Rasool, N., & Bhat, J. I. (2025). Brain tumour detection using machine and deep learning: a systematic review. Multimedia tools and applications, 84(13), 11551-11604.

[56] Zhu, C., Chai, X., Xiao, Y., Liu, X., Zhang, R., Yang, Z., & Wang, Z. (2024). Swin-Net: A swin-transformer-based network combing with multi-scale features for segmentation of breast tumor ultrasound images. Diagnostics, 14(3), 269.

[57] Omarov, B., Omarov, B., Rakhymzhanov, A., Niyazov, A., Sultan, D., & Baikuvekov, M. (2024). Development of an artificial intelligence-enabled non-invasive digital stethoscope for monitoring the heart condition of athletes in real-time. Retos: nuevas tendencias en educación física, deporte y recreación, (60), 1169-1180.

[58] Zhu, Y., Li, H., Cao, B., Huang, K., & Liu, J. (2025). A novel hybrid layer-based encoder–decoder framework for 3D segmentation in congenital heart disease. Scientific Reports, 15(1), 11891.

[59] El Badaoui, R., Bonmati Coll, E., Psarrou, A., Asaturyan, H. A., & Villarini, B. (2025). Enhanced CATBraTS for Brain Tumour Semantic Segmentation. Journal of Imaging, 11(1), 8.

[60] Mayzura, W. S., Sarno, R., Suroto, N. S., Supriyanto, M. I. A., & Sihaj, G. (2025). Automatic Interpretation of Brain Medical Images Using Hierarchical Classification and Image Captioning Model. IEEE Access.

[61] Samak, Z. A. (2025). Multi-type stroke lesion segmentation: comparison of single-stage and hierarchical approach. Medical & Biological Engineering & Computing, 63(4), 975-986.