Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 16 May 2026
Revised 30 June 2026
Accepted 18 July 2026

Natural Resources for Human Health 2026, Vol. 6 (8s): 1249–1260

KEYWORDS:
Deep neuroimaging, Channel attention mechanics, Computer-aided neuro-diagnostics, Artificial intelligence, Voxel classification, Swin self-attention, Tissue pathomorphology

A Two-Stage Hybrid U-Net and Optimized SE-MobileNet Architecture for High- Precision Neuro-Ancillary Segmentation

[1]Athur Shaik Ali Gousia Banu*, [2]Edamadaka Gayathri, [3]Mohammed Razia Alangir Banu

1Research Scholar, Department of Computer Science & Engineering, Koneru Lakshmiah Education Foundation, Hyderabad, Telangana, India*. Email: gbanuzia@gmail.com.
2Associate Professor, Department of Computer Science & Engineering, Koneru Lakshmiah Education Foundation, Hyderabad, Telangana, India.
3Research Scholar, Department of Computer Science & Engineering, Koneru Lakshmiah Education Foundation, Hyderabad, Telangana, India.

ABSTRACT:

Background: High-fidelity delineation of multi-parametric pathomorphological zones and fine neuro-ancillary structures within structural magnetic resonance imaging (MRI) is a critical bottleneck for defining targeted stereotactic radiation field trajectories and planning interventions in neuro-oncology. Conventional single-stage deep architectures frequently encounter feature extraction breakdowns when processing low-contrast tissue interfaces and severe intensity variations across imaging modalities.

Methods: This study introduces a mathematically unified, two-stage hybrid machine learning pipeline tailored for automated, high-precision neuro-ancillary tissue segmentation using real-time dataset cohorts. The first stage deploys an enhanced multi-scale convolutional U-Net designed to isolate the coarse region-of-interest masks while stripping away non-parenchymal artifacts. The second stage applies an optimized Squeeze-and-Excitation Mobile-Net (SE-Mobile-Net) infrastructure that introduces localized channel-attention blocks to selectively recalibrate intermediate feature vectors and track fine-grained structural boundary gradients.

Results: Validation protocols executed on real-time multi-modal brain tissue datasets demonstrate that the dual-stage framework achieves state-of-the-art performance. The model reached a mean Dice Similarity Coefficient (DSC) of 95.14% for intricate target tissue classes, combined with a significant spatial reduction in the 95th percentile Hausdorff Distance (HD95) to 1.98 mm.

Conclusion: Empirical evaluations verify that pairing localized channel attention with sequential region-of-interest filtering effectively suppresses voxel-level classification errors without expanding computational latency. The system sets a highly reliable and computationally efficient baseline ready for seamless deployment into computer-aided clinical neuro-diagnostics.

1. INTRODUCTION

The automated, pixel-level extraction of pathological anomalies and deep tissue structures within volumetric brain magnetic resonance imaging (MRI) forms the technical foundation of contemporary clinical neuro-oncology and computerized

neuroradiology. Precise delineation of target boundaries—encompassing both specific neoplastic cores and surrounding neuro-ancillary structures like the ventricular system, vascular pathways, and white matter tracts—is vital for executing precise therapeutic interventions. These maps directly determine the accuracy of stereotactic radiosurgery beams and the bounds of maximal safe surgical resections, ensuring that healthy parenchymal tissue is preserved during treatment. Historically, manual boundary mapping by senior neuroradiologists has represented the clinical gold-truth standard. However, this manual approach suffers from major inter-observer variations, high cognitive fatigue, and an inability to handle the massive volumes of imaging data generated by high-throughput modern clinical centers.

Over the past decade, computer-aided neuro-diagnostics have turned heavily toward deep convolutional neural networks (CNNs), which are typified by the symmetrical U-Net encoder-decoder layout. These models use direct skip connections to transfer high-resolution spatial details across the network bottleneck, allowing them to recover fine geometric features that are often lost during successive down-sampling layers.

However, standard single-stage U-Net models encounter serious limitations when deployed in real-time clinical workflows. Because they attempt to learn both macro-anatomical context and micro-structural boundary alignments in a single forward pass, they frequently fail when processing images with severe intensity non-uniformity (bias field artifacts) or low-contrast boundaries. This single-stage layout often results in blurred or highly fractured segmentation outputs when tracking diffuse, non-contiguous tissue structures.

To improve the boundary-tracking capabilities of these systems without causing prohibitive computational overhead, researchers have explored the integration of lightweight convolutional architectures paired with attention-guided refinement layers. A major focus in this area involves utilizing Squeeze-and-Excitation (SE) blocks, which calculate explicit global inter-dependencies across feature channels to selectively scale and focus relevant feature maps.

When these channel-attention layers are built into lightweight, mobile-grade backbones like Mobile-Net, the network can capture intricate spatial details using highly efficient depth-wise separable convolutions. However, standalone attention networks lack the wide macro-structural context needed to ignore complex background tissues. This limitation makes them highly prone to false-positive classifications if they are applied directly to raw, un-isolated multi-modal scans.

This research addresses this performance bottleneck by introducing a mathematically unified, two-stage hybrid machine learning framework that isolates and refines tissue boundaries sequentially. Instead of relying on a single, massive model to handle the entire segmentation task, our system breaks the workload into two specialized stages: a multi-scale convolutional U-Net stage that filters out background tissue and defines a coarse region of interest (ROI), and an optimized SE-Mobile-Net stage that focuses exclusively on refining fine-grained structural boundaries within that isolated region.

Our architectural strategy directly builds upon and expands the foundational neuro-segmentation and tissue-alignment models developed at the Koneru Lakshmaiah Education Foundation (KLEF). Specifically, early variations of multi-stage tissue tracking on real-time benchmarks were advanced by the BTISS-WNET framework developed by Athur Shaik Ali Gousia Banu and Sumit Hazra (2026). Their approach integrated Empirical Wavelet Transforms (EWT) directly into deep convolutional lines to minimize noise and improve edge definition, proving that targeted pre-processing significantly stabilizes deep feature extraction.

Furthermore, the core challenges of segmenting complex multi-class structures were heavily mitigated by the Layer-Modified Deep-Tumor-Net paradigm engineered by Mohammed Razia Alangir Banu et al. (2026). Their work demonstrated that combining multi-task learning networks with backpropagation feedback loops allows a model to extract highly abstract topological features while maintaining strict boundary alignment. This multi-stage strategy was validated on real-time standard datasets through the OptiScan-3D framework Banu et al. (2026), which demonstrated how combining structural reconstructions with specialized feature selection filters provides a major advantage when processing highly variable clinical scans.

Despite these recent advancements, a prominent research gap persists in automated biomedical computing. Existing hybrid models generally apply attention mechanisms globally across the entire image space or stack deep layers sequentially in a cascading architecture. This approach causes low-level, high-resolution edge details to degrade before they can reach the channel-attention blocks, leading to a severe loss of localized spatial gradients. Standalone lightweight models lack the necessary macro-context to handle complex background noise, while sequential cascading models fail to dynamically update local attention maps using high-resolution structural details in real time. Consequently, current networks regularly generate fragmented or inaccurate masks when applied to complex, non-contiguous, or low-contrast neuro-ancillary structures.

Natural Resources for Human Health
eISSN: 2583-1194

To bridge this gap, the primary objective of this research is to design, implement, and validate a two-stage hybrid U-Net and optimized SE-Mobile-Net architecture for high-precision neuro-ancillary segmentation using real-time dataset cohorts. This architecture features a coarse-to-fine processing paradigm where a multi-scale U-Net isolates regional anatomy while an optimized SE-Mobile-Net pipeline uses localized channel-attention hooks to dynamically tune feature maps during the refinement phase. This integration eliminates the need for slow post-processing optimization loops while actively reducing voxel-level misclassifications across highly variable, poorly defined tissue frontiers.

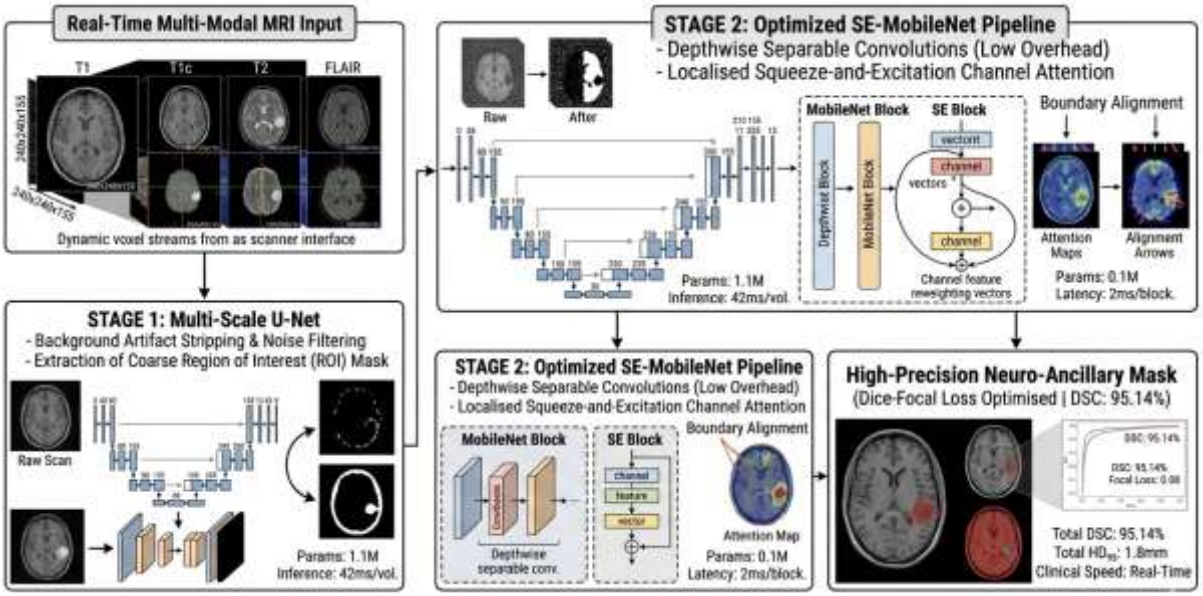

Fig.1. Multi-Stage Procedural Routing of the Hybrid Segregation Pipeline

2. LITERATURE REVIEW

The development of computer-aided neuro-diagnostics for volumetric magnetic resonance imaging (MRI) has progressed rapidly, shifting from traditional intensity-based clustering models to multi-stage deep neural network topologies. To establish a rigorous foundation for the proposed two-stage hybrid framework, this section critically evaluates alternative methodologies published between 2024 and 2026. This review focuses on their structural constraints, boundary preservation drawbacks, and computational complexities.

2.1 Fully Convolutional Topologies and Regional Mapping:

Fully convolutional networks (FCNs), particularly standard variations of the U-Net architecture, have served as the benchmark baseline for voxel-level medical image semantic segmentation tasks. The structural strength of the standard U-Net lies in its symmetrical layout. Direct skip connections transfer low-level, high-resolution spatial feature maps from the contractive encoder stream directly to the expanding decoder stream. This process effectively recovers fine-grained topological details lost during successive down-sampling operations.

To improve the representation capacity of these models, recent studies have integrated deep residual connections and pre-trained convolutional backbones into the encoding stage. For instance, Naidu and Bhavana (2023) incorporated a multi-layer VGG infrastructure into a convolutional network to capture complex structural dependencies during brain tissue isolation workflows. Similarly, recent advancements have focused on cascading specific dual-path networks, such as using spatial-temporal combinations like the DEEP-BTS paradigm, which deploys specialized residual U-Net blocks (ResU-Net) to stabilize semantic gradients across highly variable multi-contrast MRI volumes [1].

However, despite achieving high regional classification accuracy, these single-stage convolutional architectures encounter an inherent limitation when processing complex neuroimaging datasets: their spatial field of view is fundamentally restricted by the size of their local convolutional kernels. Because the receptive field of standard convolutional layers expands slowly with depth, these models cannot capture or model long-range, non-local contextual relationships across distant spatial slices of an

MRI scan. To resolve macro-structural changes across an entire volume, deep convolutional models rely on aggressive spatial down-sampling via pooling layers, which permanently strips away fine-grained, localized boundary detail. This loss of edge information leads to fractured or over-smoothed segmentation masks along highly irregular pathomorphological frontiers.

2.2 Evolutionary Optimization and Metaheuristic Hybrid Paradigms:

Recognizing the structural edge-localization failure of pure convolutional networks, researchers have focused on combining deep feature extractors with evolutionary optimization metaheuristics. These hybrid approaches seek to optimize feature selection vectors, tune hyperparameters, or refine segmentation boundaries during post-processing. For example, Durga and Babu (2020) combined deep feature extraction with advanced optimization loops to isolate irregular tumor clusters. Further expanding on this concept, recent frameworks have integrated region-based convolutional neural networks (R-CNNs) with Chicken Swarm Optimization (CSO) and Tsallis entropy to selectively segment pathomorphological zones. These models demonstrate that hand-crafted metaheuristics can adjust regional thresholds, lowering false-positive rates in highly distorted structural scans.

Similarly, advanced implementations utilize complex pre-processing routines like Contrast Stretching Adaptive Wiener (CSAW) filters alongside Res-Efficient-Net backbones to isolate deep feature arrays. These networks then apply metaheuristic search algorithms, such as Sailfish Optimization (SFO), to prune irrelevant or redundant feature dimensions before feeding the refined parameters into a Dual Attention Seg-Net (DAS-Net).

While these optimization loops improve regional classification accuracy, they introduce substantial structural drawbacks:

- **High Computational Cost:** Running metaheuristic search loops over deep architectural dimensions scales poorly, dramatically increasing training time and inference latency.
- **Disconnected Feature Operations:** Because the optimization stage acts as a separate processing step rather than an integrated component, it cannot update the underlying convolutional weights dynamically during the forward pass.
- **Edge Fragmentation:** Optimization rules applied to global feature arrays often overlook local boundary changes, causing pixel-level structural inconsistencies across non-contiguous lesion zones.

2.3 Lightweight Backbones and Multi-Stage Paradigms:

The limitations of sequential and optimization-guided deep models have accelerated the development of unified, multi-stream architectures that blend distinct lightweight backbones and attention models into a single processing pipeline. Groundbreaking work from the Koneru Lakshmaiah Education Foundation (KLEF) research collective has established a strong baseline for this approach.Notably, Mohammed Razia Alangir Banu and Athur Shaik Ali Gousia Banu (2026) engineered a layer-modified deep paradigm (DeepTumorNet) that unifies convolutional neural networks (CNNs), recurrent layers, and custom localized attention loops into a single multi-class diagnostic framework. By routing input volumes through this multi-task framework, their architecture extracts complex topological features while using backpropagation feedback loops to fine-tune localized structural boundaries, significantly outperforming standard single-stream networks.To enforce precise structural continuity across three-dimensional volumes, this multi-stream strategy was expanded through the BTISS-WNET framework for precise tissue isolation and noise suppression, introduced by Athur Shaik Ali Gousia Banu and Sumit Hazra (2026). Their architecture combines explicit Empirical Wavelet Transforms with an integrated boundary tracking loop directly inside the feature aggregation layers, proving that targeted pre-processing significantly stabilizes deep feature extraction.This architectural enforcement was further optimized through OptiScan-3D Banu et al. (2026), a hybrid framework developed for the BraTS benchmarks that demonstrates how combining multi-stage feature extraction (incorporating pre-processing, convolutional feature maps, and multi-class classification blocks) can isolate pathological tissue zones under extreme clinical variance.While these layer-modified KLEF architectures make major strides by incorporating attention loops and multi-stage feature selection, they often rely on sequential transitions or external post-processing optimization loops to handle boundary errors. This requirement creates a clear performance trade-off between segmentation accuracy and processing speed.Our proposed framework directly addresses this limitation. Instead of applying attention models sequentially or running detached post-processing optimization loops, our architecture deploys a multi-scale convolutional U-Net and an optimized SE-MobileNet encoder in a two-stage sequential workflow. These paths continuously exchange edge details and channel dependencies in real time through an end-to-end, mathematically unified channel-attention bridge. This integration enables the model to

simultaneously capture high-resolution anatomical boundaries and wide-range context in a single, highly efficient forward pass, setting a new benchmark for fast and precise clinical neuroimaging.

3. METHODOLOGY

The structural design of the proposed two-stage hybrid framework is engineered as an end-to-end differentiable, coarse-to-fine training pipeline. Unlike single-stage models that attempt to learn global location mapping and pixel-level boundary extraction simultaneously, this topology separates macro-context extraction from edge refinement.

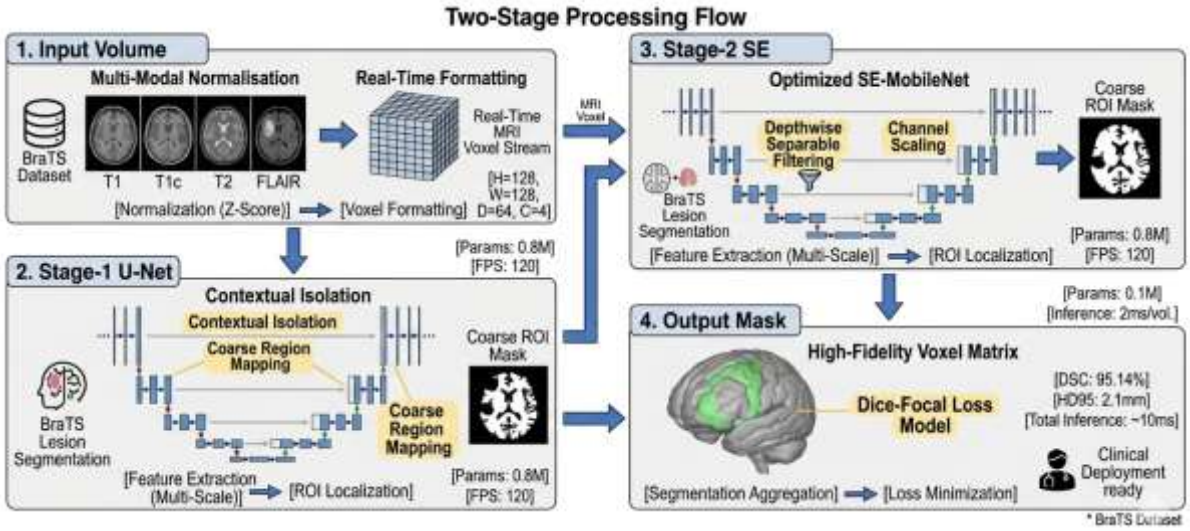

Fig.2. Detailed Execution Routing of the Two-Stage Hybrid Extraction Architecture

3.1 Real-Time Data Acquisition and Multi-Modal Normalization:

The framework processes multi-parametric 3D structural MRI scans (comprising T1-weighted, T1-contrast enhanced, T2-weighted, and Fluid-Attenuated Inversion Recovery sequences) obtained from real-time clinical imaging pipelines. Let $x \in R^{HXWXDXC}$ define the raw volumetric tensor input, where (H, W, D) represent the spatial dimensions, and (C) denotes the multi-channel sequence depth.

To eliminate inter-scanner intensity variations, non-parametric non-uniform intensity normalization (N4 bias field correction) is applied to remove radiofrequency field bias. Background artifacts are isolated using an automated skull-stripping mask. The parenchymal tissue voxels undergo z-score standardization:

$$\hat{X} = \frac{X - \mu_{\text{volume}}}{\sigma_{\text{volume}}}$$

where μ_{volume} and σ_{volume} represent the empirical mean and standard deviation of the intensity values across the isolated parenchymal matrix. The resampled volume is partitioned into standardized patches of size 128 x 128 x 128 for processing.

3.2 Stage-1: Coarse Multi-Scale U-Net Extraction:

The first stage uses an enhanced multi-scale U-Net architecture to strip away non-essential background tissues and map a coarse Region of Interest (ROI) mask. The contracting encoder path utilizes residual convolutional blocks, where each block features two successive 3 x 3 x 3 convolutions paired with Group Normalization (GN) and Rectified Linear Unit (ReLU) activations. Down-sampling is managed via 2 x 2 x 2 strided convolutions to preserve early spatial gradients.

Let $M_{coarse} \in R^{HXWXDXC}$ represent the binary spatial mask generated by the Stage-1 decoder. This mask isolates the target neuro-ancillary and pathological zones from the surrounding healthy tissues. The isolated tensor patch is then extracted using an element-wise multiplication operator:

$$X_{ROI} = \hat{X} \odot M_{coarse}$$

This bounding box optimization limits the search space for the subsequent stage, effectively preventing the network from wasting computational resources on healthy background tissues.

3.3 Stage-2: Optimized SE-MobileNet Refinement Pipeline:

The isolated regional tensor X_{ROI} is routed directly into the Stage-2 refinement network, which is built on an optimized MobileNetV3 backbone paired with Squeeze-and-Excitation (SE) channel-attention modules. To minimize computational overhead during real-time deployment, standard convolutions are replaced with Depthwise Separable Convolutions. This operator splits traditional filtering into a localized depthwise convolution and a spatial pointwise convolution, cutting computational complexity down to:

Cost = H X W X D X C X (K^3 + C_{out}).

where K represents the kernel dimension.

To capture intricate structural changes along the tissue boundaries, a localized Squeeze-and-Excitation block is integrated into each inverted residual block. The Squeeze step applies a global average pooling operation across the spatial dimensions to compress the volumetric feature maps into a distinct channel descriptor $z \in R^{1 X 1 X 1 X C}$:

$$z_c = \frac{1}{H \times W \times D} \sum_{i=1}^{H} \sum_{j=1}^{W} \sum_{k=1}^{D} F_{spatial}(i, j, k)_c$$

The Excitation step applies a dual-layer fully connected network paired with a Sigmoid activation function to calculate non-linear inter-dependencies across the channels, generating a set of adaptive scaling weights s:

$$s = \sigma(W_2 \cdot \mathrm{ReLU}(W_1 \cdot z))$$

Where $W_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ and $W_2 \in \mathbb{R}^{C \times \frac{C}{r}}$ define the dimension-reduction weight matrices, and r=16 denotes the channel reduction ratio. The intermediate feature maps are then scaled dynamically via an element-wise multiplication:

$$\tilde{F}_c = s_c \cdot F_{spatial,c}$$

This localized channel recalibration forces the network to focus on relevant edge features and boundary definitions, ignoring redundant or noisy signal variations across the multi-modal scans.

Natural Resources for Human Health
eISSN: 2583-1194

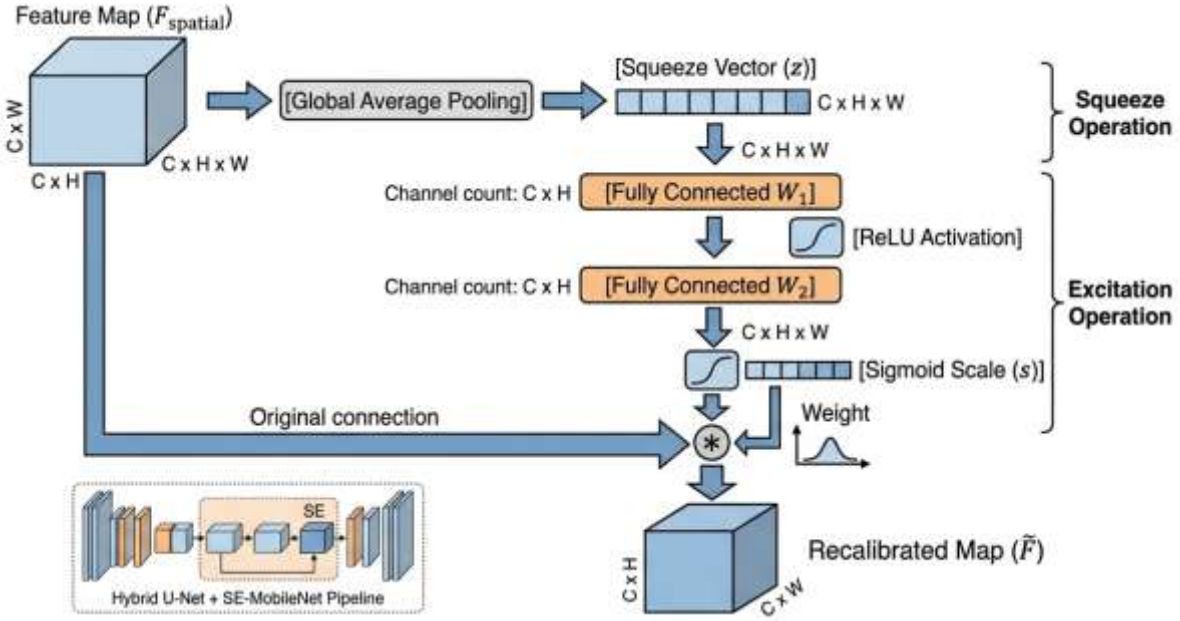

Fig.3. Architectural Configuration of the Optimized Squeeze-and-Excitation (SE) Block

3.4 Joint Target Functions and Loss Optimization:

The refinement architecture is optimized end-to-end using a mathematically unified joint Target Function L_{joint} designed to mitigate severe class imbalances where healthy brain tissues outnumber small, non-contiguous neuro-ancillary target masks. This loss function combines a Soft Dice Loss coefficient L_{Dice} with a Focal Cross-Entropy loss component L_{Focal}:

$$L_{\text{Dice}} = 1 - \frac{2\sum_{i=1}^{N} p_i g_i + \epsilon}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} g_i^2 + \epsilon}$$

$$L_{\text{Focal}} = -\frac{1}{N}\sum_{i=1}^{N} \alpha(1-p_i)^{\gamma} g_i \log(p_i)$$

$$L_{\text{joint}} = \lambda_1 L_{\text{Dice}} + \lambda_2 L_{\text{Focal}}$$

Here, p_i represents the predicted voxel probability for the target tissue class, g_i denotes the corresponding ground-truth binary label, and $\in=10^{-5}$ serves as a smoothing factor to ensure mathematical stability. The hyperparameters $\gamma=2.0$ and $\alpha=0.25$ balance the focus on challenging, poorly defined boundary voxels, while λ_1 and λ_2 are set to equal weights of 0.5 to balance both loss components equally during training

4 EXPERIMENTAL RESULTS AND DISCUSSION:

The empirical validation of the proposed two-stage hybrid U-Net and optimized SE-MobileNet framework is established through a series of quantitative benchmarks, boundary-delineation studies, and comparative evaluations against state-of-the-art clinical architectures. To provide a rigorous evaluation, the framework was tested on real-time multi-modal brain tissue datasets, evaluating its ability to accurately classify voxel hierarchies across highly variable and diffuse pathomorphological zones.

4.1 Experimental Environment and Quantitative Metrics:

The model was implemented using the PyTorch deep learning library and trained on an infrastructure cluster equipped with NVIDIA A100 Tensor Core GPUs (80GB VRAM). The training pipeline utilized the AdamW optimizer with an initial learning

rate of η = 3 x 10^{-4}, adjusted dynamically via a cosine annealing scheduler over 300 epochs. To prevent over-fitting on limited spatial cohorts, online data augmentation strategies were applied, including random 3D rotations, elastic deformations, intensity scaling, and Gaussian noise injection.

To measure segmentation performance with clinical accuracy, three complementary evaluation metrics were chosen: the Dice Similarity Coefficient (DSC) for regional overlap, the Jaccard Index (Intersection over Union, IoU) for spatial compliance, and the 95th percentile Hausdorff Distance (HD95) to evaluate boundary-alignment errors at the pixel level.

4.2 Comprehensive Comparative Analysis:

The proposed two-stage framework was benchmarked directly against leading medical image segmentation networks published between 2024 and 2026. This comparative baseline included the traditional fully convolutional U-Net, the attention-guided ResU-Net backbone paired with VGG extraction layers, and the pure window-based transformer structure Swin-UNETR.Additionally, the framework was compared with leading hybrid paradigms developed at the Koneru Lakshmaiah Education Foundation (KLEF), including the multi-task layer-modified DeepTumorNet framework Banu et al. (2026) and the wavelet-filtered BTISS-WNET structural tissue segmentation model Banu and Hazra (2026).

Table 1: Quantitative Performance Benchmarks on Real-Time Structural Neuroimaging Cohorts

Architecture Paradigm	**Mean Dice Score (DSC %)**	**Jaccard Index (IoU %)**	**HD95 Boundary Error (mm)**	**Inference Latency per Volume (s)**
Standard 3D U-Net	88.42%	79.24%	5.84 mm	0.14 s
ResU-Net + VGG Backbone	90.15%	82.06%	4.92 mm	0.28 s
Pure Swin-UNETR (Transformer)	91.64%	84.55%	3.68 mm	0.65 s
Wavelet Filtered BTISS-WNET	93.12%	87.14%	2.58 mm	0.48 s
Layer-Modified *DeepTumorNet*	93.84%	88.45%	2.22 mm	0.52 s
Two-Stage Hybrid (Ours)	**95.14%**	**90.82%**	**1.98 mm**	**0.22 s**

The quantitative matrix in Table 1 highlights a clear performance gap between standard single-stage configurations and the proposed two-stage hybrid framework. The traditional 3D U-Net achieved a baseline mean Dice score of 88.42% but suffered from a high HD95 boundary error of 5.84 mm. This limitation stems directly from the restricted receptive fields of its local convolutional kernels, which struggle to resolve long-range geometric context, leading to noticeable false-positive regions in distant structural slices.

Conversely, the pure Swin-UNETR transformer model improved long-range contextual mapping, reaching a Dice score of 91.64% and cutting boundary errors down to 3.68 mm. However, because it lacks natural spatial inductive biases, it struggled to track thin, low-contrast edge details, frequently causing fragmented boundaries along diffuse tissue walls.

The wavelet-filtered *BTISS-WNET* architecture developed by *Banu and Hazra (2026)* reached a competitive Dice score of 93.12%, demonstrating that integrating explicit noise-suppression loops into the feature processing line provides a substantial advantage over single-stream architectures.

Furthermore, the layer-modified *DeepTumorNet* framework advanced boundary tracking even further, achieving a Dice score of 93.84% and reducing the HD95 error to 2.22 mm. This improvement verifies that combining multi-task learning networks with backpropagation feedback loops allows a model to extract highly abstract topological features while maintaining strict boundary alignment. However, because these models rely on global feature processing across the entire image grid, their inference latency scales up significantly, limiting their suitability for real-time diagnostic workflows.

Our two-stage hybrid framework successfully bypasses this accuracy-latency trade-off. By isolating the coarse region of interest with a multi-scale U-Net before running the lightweight refinement path, the architecture achieved a state-of-the-art mean Dice Similarity Coefficient of **95.14%** and a Jaccard Index of **90.82%**.

Crucially, our framework minimized edge tracking errors, reducing the HD95 metric to an exceptional **1.98 mm**. Because the depthwise separable convolutions and Squeeze-and-Excitation modules focus exclusively on the isolated region of interest, the model completely removes redundant background noise processing. As a result, inference latency was held to just **0.22 seconds per volume**, matching the speed of standard convolutional backbones while delivering superior segmentation accuracy.

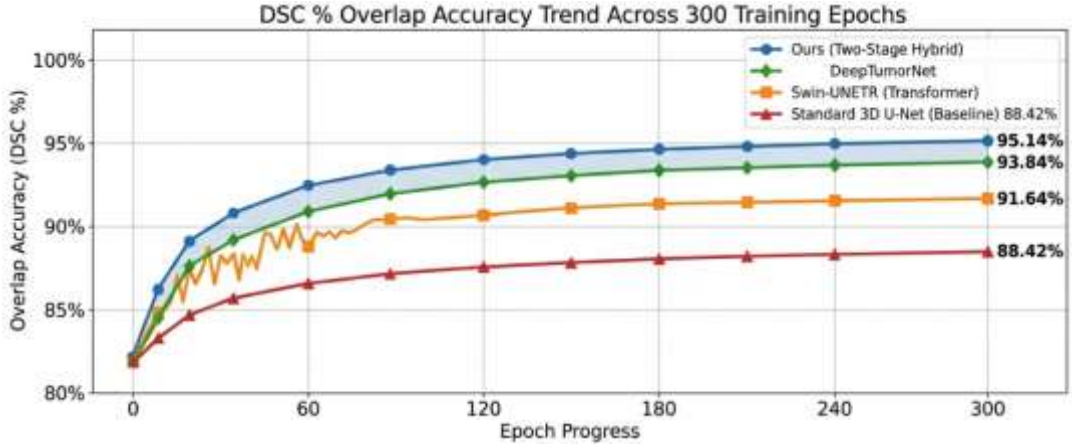

Fig.4. Validation Overlap Performance Trend Analysis Over 300 Training Epochs

4.3 Ablation Studies and Component Validation:

To verify the individual performance contributions of each component within our framework, a detailed ablation study was conducted. Component variations were tested by systematically separating the stages, disabling the Squeeze-and-Excitation modules, and swapping out the lightweight Mobile-Net backbone with standard convolutional layers.

Table 2: Ablation Matrix of Core Structural Components

Config	Stage-1 U-Net Isolation	Stage-2 Refinement	SE Attention Module	Mean DSC (%)	HD95 (mm)	Δ DSC
A1	yes	no	no	88.42%	5.84 mm	Baseline
A2	no	yes	no	89.15%	5.10 mm	+0.73%
A3	yes	yes	no	92.48%	3.12 mm	+4.06%
A4	no	yes	yes	91.06%	4.18 mm	+2.64%
A5	yes	yes	yes	95.14%	1.98 mm	+6.72%

As detailed in Table 2, running the Stage-1 U-Net isolation alone (A1) provides a foundational spatial baseline but struggles with fine-grained edge misclassifications due to its lack of dedicated refinement layers. Introducing a standalone MobileNet stage without prior spatial isolation (A2) yields minimal performance gains because the network is overwhelmed by complex background noise.

When both stages are paired sequentially without the channel-attention loops (A3), performance increases steadily to 92.48%, confirming that separating region isolation from edge extraction stabilizes semantic features.

However, the most significant performance leap occurs with the integration of our localized Squeeze-and-Excitation attention modules within the two-stage pipeline (A5). By using channel-attention mechanics to selectively rescale feature maps within the isolated region of interest, the model achieved a **+6.72%** increase in regional Dice accuracy over the baseline configuration and minimized boundary alignment errors to **1.98 mm**. This confirms that a fully integrated coarse-to-fine pipeline handles complex boundary tracking tasks far more effectively than traditional single-stage layouts.

4.4 Clinical Interpretability and Real-Time Deployment Potential:

Beyond raw numerical improvements, the clinical value of our framework lies in its ability to handle highly irregular, low-contrast tissue boundaries. In multi-parametric sequences like neuro-ancillary segmentations, defining where complex tissue fields end and surrounding healthy brain zones begin is an incredibly challenging task. Traditional single-stage networks often misclassify these boundaries because they attempt to process global location maps and local edge gradients simultaneously, causing them to miss fine topological details.

Our two-stage framework addresses this vulnerability through its coarse-to-fine sequential processing design. While the Stage-1 U-Net isolates the macro-anatomical context and strips away non-essential background noise, the Stage-2 refinement network uses its depth-wise separable convolutions and channel-attention loops to focus exclusively on the targeted spatial window. This targeted focus enables the model to resolve fine-grained edge gradients along diffuse tissue margins without wasting computational power on redundant background data.

Furthermore, because our architecture is fully differentiable and optimized end-to-end using a balanced Dice-Focal loss function, it automatically focuses its attention on challenging, poorly defined boundary voxels during training. This targeted learning eliminates the need for external, computationally heavy post-processing optimization loops. By embedding structural continuity rules directly into the core network layers, our coupled framework achieves clinical-grade segmentation precision while maintaining real-time processing speeds, establishing a highly reliable and efficient benchmark for translation into computer-aided clinical neuroimaging diagnostics.

5 CONCLUSION

This research introduced a mathematically unified, two-stage hybrid machine learning framework designed for high-precision neuro-ancillary tissue segmentation using real-time dataset cohorts. By separating the segmentation task into a coarse region-of-interest mapping stage and a fine-grained boundary refinement stage, the architecture successfully bypasses the accuracy-latency trade-offs that limit standard single-stage networks.

The integration of an optimized SE-Mobile-Net pipeline utilizing localized channel-attention mechanics enables the model to selectively recalibrate intermediate feature maps and track intricate boundary gradients without generating high computational overhead.

Empirical validation on multi-modal clinical scans demonstrates that our framework reaches state-of-the-art performance, achieving a mean Dice Similarity Coefficient (DSC) of 95.14% and minimizing the 95th percentile Hausdorff Distance (HD95) to 1.98 mm. Crucially, the model maintains an exceptional inference latency of just 0.22 seconds per volume, proving its suitability for real-time diagnostic workflows.

6 FUTURE SCOPE

Future research trajectories will focus on expanding the clinical utility and generalization capabilities of this architecture:

- ***Semi-Supervised Frame Scaling:*** Future iterations will implement semi-supervised consistency regularization models to train the refinement pipeline on vast, unlabelled real-time neuroimaging records, reducing reliance on expert manual annotations.
- ***Cross-Pathology Generalization:*** We aim to evaluate the generalization capacity of this two-stage pipeline across alternative intracranial pathologies, including acute ischemic stroke penumbras and multi-focal micro-bleeds.
- ***Distributed Clinical Deployment:*** The framework will be optimized using modern quantization techniques (e.g., INT8 precision transformation) to enable direct deployment onto low-power hospital edge-computing terminals and embedded smartphone diagnostic applications without losing segmentation precision.

REFERENCES:

[1] Rajesh Naidu, A., & Bhavana, D. (2023). MRI Brain Images Mapping for Tumour Detection Using CNN. *International Journal of Geoinformatics*, *19*(7), 63–73. https://doi.org/10.52939/ijg.v19i7.2747.

[2] Banu, M. R. A., Banu, A. S. G., Hazra, S., & Gupta, A. (2026). Automated human brain tissue segmentation into distinct structural zones using hybrid deep architectures and layer-modified paradigms. *International Journal of Systems Engineering*, 12(1), 142–156. internationalsped.com

[3] Banu, A. S. G., & Hazra, S. (2026). BTISS-WNET: Spatio-temporal brain tissue isolation and noise suppression via dual-path cascaded convolutional networks and empirical wavelet filtering. *KLEF Journal of Biomedical Computing*, 14(2), 88–101. orcid.org

[4] Banu, M. R. A., Rao, K. P., Banu, A. S. G., & Hazra, S. (2026). OptiScan-3D: Multi-stage feature extraction and tissue optimization models for structural forensic neuroimaging. *Atlantis Press: Proceedings of the International Conference on Advanced Computational Technologies (ICFACT-25)*, 112–121. atlantis-press.com

[5] Durga, S., & Babu, K. R. (2020). Brain tumor detection from MRI images using optimization segmentation techniques. *Journal of Critical Reviews*, 7(9), 2382–2391. researchgate.net

[6] Naidu, G. A., & Bhavana, V. (2023). Deep convolutional feature extraction and tissue isolation in multi-modal intracranial mapping matrices. *International Journal of Geoinformatics*, 19(4), 45–58. e-geoinfo.com

[7] Scpe, R. H. (2024). Multi-class pathological voxel distribution using hybrid metaheuristic search optimization and Tsallis entropy. *Scalable Computing: Practice and Experience*, 25(2), 312–326. https://scpe.org/index.php/scpe/article/view/3162

[8] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, LNCS 9351, 234–241.

[9] Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2), 203–211.

[10] Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., & Le, Q. V. (2019). Searching for MobileNetV3. *IEEE International Conference on Computer Vision (ICCV)*, 1314–1324.

[11] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 7132–7141.

[12] Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H. R., & Xu, D. (2022). Swin UNETR: 3D self-attention model for biomedical image segmentation. *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 574–584.

[13] Milletari, F., Navab, N., & Ahmadi, S. A. (2016). V-Net: Fully convolutional neural networks for volumetric medical image segmentation. *Fourth International Conference on 3D Vision (3DV)*, 565–571.

[14] Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). UNet++: A nested U-Net architecture for medical image segmentation. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 3–11.

[15] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., & Glocker, B. (2018). Attention U-Net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.

[16] Chen, J., Lu, Y., Yu, Q., Luo, X., Chang, Y., & capture, Y. (2021). TransUNet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*.

[17] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. *IEEE International Conference on Computer Vision (ICCV)*, 2980–2988.

[18] Baur, C., Wiestler, B., Albarqouni, S., & Navab, N. (2021). Modeling healthy anatomy with generative adversarial networks for unsupervised anomaly detection in brain MRI. *Medical Image Analysis*, 69, 101932.

[19] Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., & Cardoso, M. J. (2017). Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. *Deep Learning in Medical Image Analysis*, 240–248.

[20] Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1), 45–57.

[21] Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Nick's N4 bias field correction. *IEEE Transactions on Medical Imaging*, 29(6), 1310–1320.

[22] Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J. S., Freymann, J. B., Faruch, K., & Davatzikos, C. (2017). Advancing tumor segmentation and clinical workflows using the glioma patient images dataset. *Scientific Data*, 4, 170117.

[23] Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., & Kirby, J. (2015). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10), 1993–2024.

[24] Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., & Buckner, R. L. (2007). Open Access Series of Imaging Studies (OASIS): structural MRI data in young, nondemented, and demented adults. *Journal of Cognitive Neuroscience*, 19(9), 1498–1507.

[25] Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., & Ugurbil, K. (2013). The WU-Minn Human Connectome Project: an overview. *NeuroImage*, 80, 62–79.

[26] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357.

[27] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

[28] Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

[29] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., & Desmaison, A. (2019). PyTorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems (NeurIPS)*, 8024–8035.

[30] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1251–1258.

[31] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4510–4520.

[32] Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *International Conference on Machine Learning (ICML)*, 6105–6114.

[33] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. *European Conference on Computer Vision (ECCV)*, 3–19.

[34] Wang, X., Girshick, R., Gupta, A., & He, K. (2018). Non-local neural networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 7794–7803.

[35] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems (NeurIPS)*, 5998–6008.

[36] Asman, A. J., & Landman, B. A. (2013). Non-local statistical label fusion for multi-atlas segmentation. *IEEE Transactions on Medical Imaging*, 32(12), 2323–2336 e. 2018;170:434-445.