Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 3, July 2026, pp: 967-982 e-ISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

A Cross-Scale Spatial–Channel Attention Inception Network for Efficient Medical Image Segmentation

B. Krishnakumar[1], P. Nisha[2], Sri Laxmi Kuna[3], K. Venu[4], R. Evance Leethial[5], Kunchanapalli Rama Krishna[6]

[1] School of Computing, SASTRA Deemed University, Tamil Nadu, India.
[2] Department of Computer Science and Engineering, Dr. N.G.P. Institute of Technology, Coimbatore, Tamilnadu, India.
[3] Department of Computer Science and Engineering (Data Science) CVR College of Engineering, Hyderabad, India.
[4] Department of Computer Science and Engineering, Kongu Engineering College, Perundurai, India.
[5] Department of Computer Science and Engineering at Nehru Institute of Technology, Coimbatore, India
[6] Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Andhra Pradesh, India

Corresponding author: B. Krishnakumar (e-mail: krishnakumarpri@gmail.com), **Author(s) Email**: P.Nisha (e-mail: nisha.p@drngpit.ac.in), Sri Laxmi Kuna (e-mail: drsrilaxmi2019@gmail.com), K. Venu (e-mail: venu.kalaimagal@gmail.com), Evance Leethial (e-mail: v.evance08@gmail.com), Kunchanapalli Rama Krishna (e-mail: tenalirama@kluniversity.in).

Abstract **Medical image segmentation plays a crucial role in modern computerized diagnosis, as accurate delineation of anatomical structures directly impacts clinical decision-making and treatment planning. However, segmenting anatomically complex regions at a fine-grained level remains challenging, especially when computational efficiency is a key requirement. To address these challenges, the authors propose a novel, lightweight medical image segmentation framework, CSA-IncepLiteNet, designed to achieve high segmentation accuracy without imposing a significant computational burden. The CSA-IncepLiteNet architecture integrates two key innovations: cross-scale feature extraction and unified spatial channel attention learning. Central to this framework is the newly introduced Cross-Scale InceptionLite module, which efficiently captures multi-scale contextual information. This module is built using depth-wise separable convolutions and point-wise convolutions, enabling effective feature extraction while significantly reducing the number of trainable parameters. By learning features across multiple spatial scales, the network can better represent anatomically complex structures present in medical images. In addition, the authors propose a Cross-Scale Spatial Channel Attention (CSA) module that jointly models spatial saliency and channel-wise interdependencies within a unified attention-learning paradigm. This dual attention mechanism allows the network to focus on the most informative regions and feature channels simultaneously, leading to improved segmentation precision. The performance of CSA-IncepLiteNet was evaluated on the BUSI breast ultrasound dataset and multiple CT image modality-based datasets. Experimental results demonstrate that the proposed framework consistently outperforms existing state-of-the-art methods across all evaluated datasets. Notably, CSA-IncepLiteNet achieves an accuracy of 92.1% and a Dice coefficient of 82.94% on the BUSI dataset, while utilizing over 26 million fewer parameters than a conventional U-Net. These results highlight the model's effectiveness, robustness, and suitability for resource-constrained medical imaging applications.**

Keywords **Medical image segmentation, spatial–channel attention, cross-scale feature learning, lightweight deep learning, encoder–decoder networks**

I. Introduction

Medical image segmentation plays a crucial role in computer-assisted diagnosis (CAD) systems and image-guided therapy, serving as a fundamental step in modern clinical decision-making. It enables the precise delineation and characterization of anatomical

Manuscript received 15 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 1 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1550

structures, pathological regions, and abnormalities, thereby facilitating accurate disease identification and treatment planning [1]. High-precision segmentation is essential not only for diagnosis but also for preoperative planning, surgical navigation, and post-treatment evaluation across a wide range of medical imaging modalities, including ultrasound, computed tomography (CT), magnetic resonance imaging (MRI), dermatological imaging, and endoscopy. Despite its importance, medical image segmentation remains a challenging task due to inherent limitations in imaging data, such as low contrast between tissues, presence of noise, intensity inhomogeneity, blurred or fuzzy boundaries, and significant inter-patient variability in anatomical structures and pathological manifestations [2].

In recent years, deep learning-based encoder–decoder architectures, particularly U-shaped networks such as U-Net and its variants, have emerged as the dominant paradigm for medical image segmentation. These architectures effectively capture both local and global contextual information through hierarchical feature extraction and reconstruction [3]. To further enhance segmentation performance, attention mechanisms have been integrated into these networks. Spatial attention mechanisms allow the model to focus on clinically relevant regions within the image, while channel attention mechanisms emphasize informative feature maps and suppress irrelevant ones [4]. However, in most existing approaches, spatial and channel attention modules are applied separately or sequentially, limiting their ability to capture comprehensive cross-dimensional feature relationships. Moreover, many state-of-the-art models are computationally intensive and memory-demanding, making them less suitable for real-time clinical applications, especially in resource-constrained healthcare environments [5].

Another major challenge in medical image segmentation is the handling of multi-scale contextual information. Lesions, tumors, and anatomical structures vary significantly in size, shape, and appearance across different imaging modalities and patients [6]. Standard convolutional layers have a fixed receptive field, which restricts their ability to capture both fine-grained details and global contextual patterns simultaneously. Although several multi-scale strategies, such as dilated convolutions, pyramid pooling, and multi-branch architectures, have been proposed to address this issue, they often introduce additional computational complexity and increase model size [7].

These limitations highlight a clear research gap: there is a need for a segmentation framework that efficiently represents multi-scale features and simultaneously models cross-dimensional (spatial and channel) attention in a unified manner, while maintaining low computational complexity [8]. To address this gap, this work proposes CSA-IncepLiteNet, a lightweight medical image segmentation framework that integrates cross-scale feature decomposition with unified spatial–channel attention learning. The proposed model aims to achieve high segmentation accuracy while reducing the computational burden, making it more suitable for practical clinical deployment [9]. Additionally, CSA-IncepLiteNet is designed to enhance boundary delineation and improve feature discrimination, which are particularly challenging in medical images due to blurred edges, overlapping tissues, and heterogeneous pathological patterns [10]. The main contributions of the proposed work are listed below.

a) A novel Cross-Scale Inception Lite block is proposed, which efficiently leverages depth-wise separable and pointwise convolutions in modeling multi-scale context while maintaining a low parameter count.
b) A unified Cross-Scale Spatial-Channel Attention (CSA) module is introduced to jointly model spatial saliency and inter-channel relationships, enabling rich cross-dimensional feature interactions.
c) An efficient end-to-end segmentation framework is built for an accurate prediction and returns substantial computational gains on heavy backbones.
d) Extensive experiments are carried out on BUSI dataset and consistently report improvements over the current leading methods at a lightweight model cost.

The rest of the paper is organized as follows. In section II, the literature review and research gap are identified. In Section III, a CSA-IncepLiteNet architecture is proposed with primary building blocks. Section IV discusses the experimental setup and evaluation metrics, and demonstrates the results, with a few discussions and comparisons. Finally, in Section VI, a summary of the paper, as well as avenues for future research, is given.

II. State-of-the-Art Techniques

Ilesanmi et al. (2021) [6] reviewed pipelines for breast ultrasound segmentation and classification, covering both traditional and deep learning approaches. They outlined common datasets, common evaluation practices, and common failure modes, especially those involving speckle noise and low-contrast boundaries. The strong points of this review are solid problem framing and a clear taxonomy. However, they do mention one important weakness: many methods from the review era failed to demonstrate robust cross-dataset validation or deployment-oriented efficiency [6]. Erin (2025) proposed a hybrid CNN–Transformer

Manuscript received 15 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 1 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1550

approach for BUS detection and segmentation. It blends a local CNN feature with the global context of transformers. This captures long-range dependencies but has a downside [7].

Bruno et al. (2025) propose a two-step approach to segment suspicious regions first and classify it into benign or malignant with backbone networks. However, there is the risk of propagating errors-segmentation mistakes might hurt classification performance-and usually, such a pipeline tends to be heavier due to two-stage inference [8].
Wu et al. (2024) proposed MFMSNet, a CNN–Transformer hybrid utilizing interactions at multiple frequencies and scales, including an octave-style feature decomposition that is frequency-aware to better handling of ultrasound texture and noise. But the hybrid designs can increase the model complexity and memory footprint [9]. Abuowaida et al. (2025) proposed UltraSegNet, a hybrid framework with a powerful backbone along with attention and fusion strategies to handle speckle noise and ambiguity around boundaries. The strengths are improved boundary-focused segmentation and workflow-oriented motivation. Limitations could be that increased backbones and fusion modules may lead to higher parameters and computation [10]. Guo et al. (2025) developed MSRA-Net, a multi-scale, region-aware segmentation network emphasizing ROI-focused learning and scale diversity. This has its advantage in better localization due to region awareness. However, region and scale modules may require careful tuning and might not generalize uniformly across different modalities without calibration [11].

Bian et al. (2025) proposed a novel segmentation framework, ThreeF-Net, that incorporates fine-grained, multi-level feature fusion to enhance detail preservation in medical image segmentation. The model is designed to effectively capture both low-level spatial details and high-level semantic information by integrating features from different stages of the encoder–decoder architecture. This multi-level fusion strategy enables improved delineation of lesion boundaries and better completeness in segmenting pathological regions, particularly in complex medical images where structures may be subtle or irregular. As a result, ThreeF-Net demonstrates superior performance in preserving intricate details and reducing boundary leakage [12]. Aumente Maestro et al. (2025) have presented a multi-task framework that combines segmentation and classification tasks in a way that takes advantage of synergy between the two. The advantage lies in a potential increase in performance and improved data efficiency. However, it should be noted that balancing task losses in this manner is difficult and depends on curated sets and consistent labeling [13].

A hybrid architecture based on a combination of the Swin Transformer model and concepts related to EfficientNet models, using efficient attention for segmentation and classification tasks with multiple ultrasound modalities, is discussed in Nissar (2025). The benefit of this architecture is its effectiveness for global context modeling. The drawback is that transformer parts are still very heavy in terms of the computations used [14]. Jiang et al. (2026) propose an approach to improve tumor classification using a segmentation knowledge-driven global and local attention mechanism. The advantages include better explainability, as the classification is associated with the lesion's structure. However, the limitation is the dependency on good lesion segmentation and the varying sensitivity to domain shifts of multi-stage

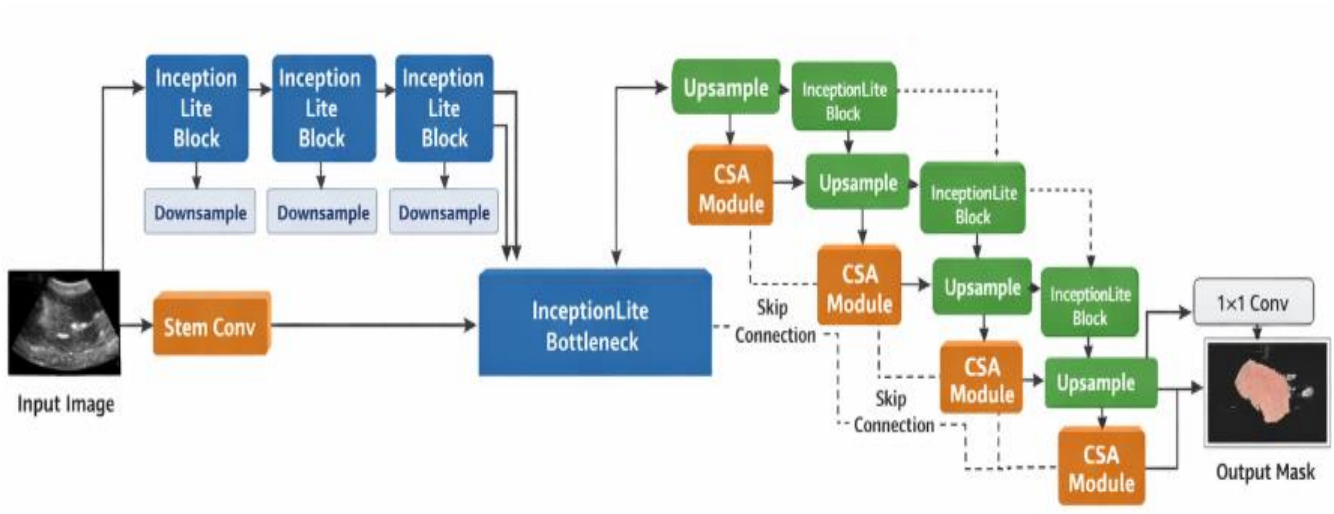


Fig. 1. An overview of the CSA-IncepliteNet

Manuscript received 15 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 1 June 2026
Digital Object Identifier (DOI): https://doi.org/10.35882/jeeemi.v8i3.1550

knowledge transfer [15]. A cross-scale attention-based U-Net was proposed by Wang et al. (2025) to facilitate interactions among features in BUS images for segmentation. The advantage of this approach is to overcome scale variations and ambiguity in boundary definitions. However, it might lead to implementation overhead, and a light-weight implementation is essential [16].

III. Proposed Work

The proposed CSA-IncepLiteNet focuses on efficient medical image segmentation by presenting the entire architecture and supporting it with complexity analysis.

A. Overall Framework Overview

Before training the proposed CSA-IncepLiteNet, a number of preprocessing steps are performed to enhance data consistency and stability in the model. First, we resized all the ultrasound images to a uniform resolution of 256x256 pixels, ensuring that the network receives uniform data in terms of resolution. Second, we normalized pixel intensities to the range [0, 1] using min-max normalization to keep gradients stable during training. Third, to enhance the generalization of the network and prevent overfitting, we performed data augmentation on the training images, including random horizontal flipping, small-angle rotation, and scaling of pixel intensity. The scaling of pixel intensity was performed to mimic the effect of brightness and contrast in ultrasound imaging. The combined effect of these steps helps to reduce noise in the ultrasound imaging data, resulting in improved performance of the proposed network. CSA-IncepLiteNet employs a U-Shape design tailored to achieve a fine balance between segmentation fidelity and computational feasibility. The encoder section extracts hierarchical features from the medical image. The decoder section creates a precise and high-resolution segmentation image by performing a series of upsampling and skip connection techniques. To enhance the discriminability of features and boundaries being used, two major concepts have been employed within CSA-IncepLiteNet [17], [18]. It contains Cross-Scale InceptionLite and Cross-Scale Spatial-Channel Attention.

Table 1. Hyperparameter Settings

Parameter	Value
Optimizer	Adam
Initial learning rate	1×10^{-4}
Batch size	8
Number of epochs	25
Weight decay	1×10^{-5}
Loss function	Dice + BCE
Input image size	256 × 256
Attention kernel size	3 × 3
Activation function	ReLU
Output activation	Sigmoid

Within the encoder, InceptionLite modules collect multi-scale context information via lightweight convolutional layers. Conversely, in the decoder, the CSA module refines the fused features by considering both the spatial importance and the channel relationships. Finally, skip connections preserve fine spatial details and allow the attention mechanism to ignore potentially irrelevant background information.

The structure of the CSA-IncepLiteNet model in the context of medical image segmentation can be depicted as in the following Fig. 1. The entire process begins with a light stem convolution that extracts basic features from the input image. The following Cross-Scale InceptionLite blocks are responsible for digging deeper into the data, extracting a hierarchical, multi-scale representation while substantially down-sampling the features in the process [19]. The bottleneck of the network consists of a more complex InceptionLite that captures contextual information in relation to making future decisions. Finally, in the decoder, the features are gradually upscaled and combined with their corresponding features in the encoder using skip connections, while finally passing through the CS Spatial-Channel Attention blocks, resulting in the accurate pixel-wise segmentation using a 1x1 convolution [20]. Table 1 lists the hyperparameters used across all experiments. The learning rate was identified as one of the key hyperparameters in determining how well the model converges during training. If the learning rate is high, it may cause unstable training, with losses oscillating. On the other hand, if it is low, it may cause slow convergence to the optimal value. The best results were achieved with an initial learning rate of $1 * 10^{-4}$ with the Adam optimizer to have a stable training process with constant improvements in Dice scores.

B. Cross-Scale InceptionLite Block

In medical images, there exist medical structures and lesions of different scales. For an effective representation of such multi-scale features, we propose an InceptionLite block, which divides the normal convolution process into several quick paths with their own receptive field, utilizing depth wise separable convolution and pointwise convolution for effective utilization of contextual information. As depicted in Fig. 2, the Cross-Scale Spatial-Channel Attention (CSA) module consists of building blocks for discovering where things are and what features matter, leading to accurate results [29].

Let $X \in R^{H \times W \times C}$ denote the input feature map. The InceptionLite block processes X through multiple parallel convolutional branches with varying kernel sizes. The output of each branch is expressed as given in Eq.(1) [19].

Manuscript received 15 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 1 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1550

Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 3, July 2026, pp: 967-982 **e-ISSN: 2656-8632**

$$F_i = PWConv\big(DWConv_i(X)\big) \quad (1)$$

DWConv stands for the application of depthwise convolution with a specific size provided, which is then performed on the i-th branch, followed by the application of PWConv, which stands for combining all the available information through a pointwise convolution process [22], [30]. Finally, after executing all the branches, the data is concatenated to give the final feature representation as in Eq.(2) [19].

$$F = Concat(F1, F2, \dots, Fn) \quad (2)$$

This structure allows the network to understand not only local information but also global information, and it reduces parameters and computations compared to typical convolutional blocks [31].

C. Cross-Scale Spatial–Channel Attention (CSA) Module

Even though such an approach greatly increases feature extraction in context, there is still no immediate guidance of the network towards meaningful clinical regions [32]. To bridge this need, we introduce our novel proposal of a unified Cross-Scale Spatial-Channel Attention (CSA) module, which, instead of using sequential spatial-channel attention, considers all aspects simultaneously [33]. For a feature map F, spatial attention can be calculated by performing a pooling operation on the channel dimension with the aid of average and max pooling operations to create a spatial descriptor indicating salient areas as given in Eq.(3) [27], [28].

$$A_s = \sigma\big(f_s([AvgPool(F); MaxPool(F)])\big) \quad (3)$$

In this configuration, f_s denotes a convolutional transformation, and σ stands for the sigmoid function. As a result, spatial attention assigns higher weights to those spatial locations that provide more informative cues [27]. Meanwhile, channel attention can be understood as derived from pooling the spatial details to emphasize the most informative feature channels as given in Eq.(4) [29].

$$A_c = \sigma\Big(f_c\big(GAP(F)\big)\Big) \quad (4)$$

where GAP denotes global average pooling and f_c represents a channel-wise transformation. The final attention-refined feature map is obtained by jointly applying spatial and channel attention as given in Eq.(5) [29].

$$F' = As \odot (Ac \odot F) \quad (5)$$

$\odot$ denotes element-wise multiplication. The unified approach within the network lets it selectively emphasize the most informative regions and discriminative feature channels while boosting the accuracy of segmentation and making the boundaries more consistent [28].

As shown, the process starts with an input feature map as input, where spatial cues are first identified using average and max pooling across the channel axis, highlighting particular features with each of these core cues [34]. These two spatial cues are then fused as an attention map using light convolution and a sigmoid function [29]. The channel attention, on the other hand, results from using spatial context cues captured globally using global average pooling, which is then

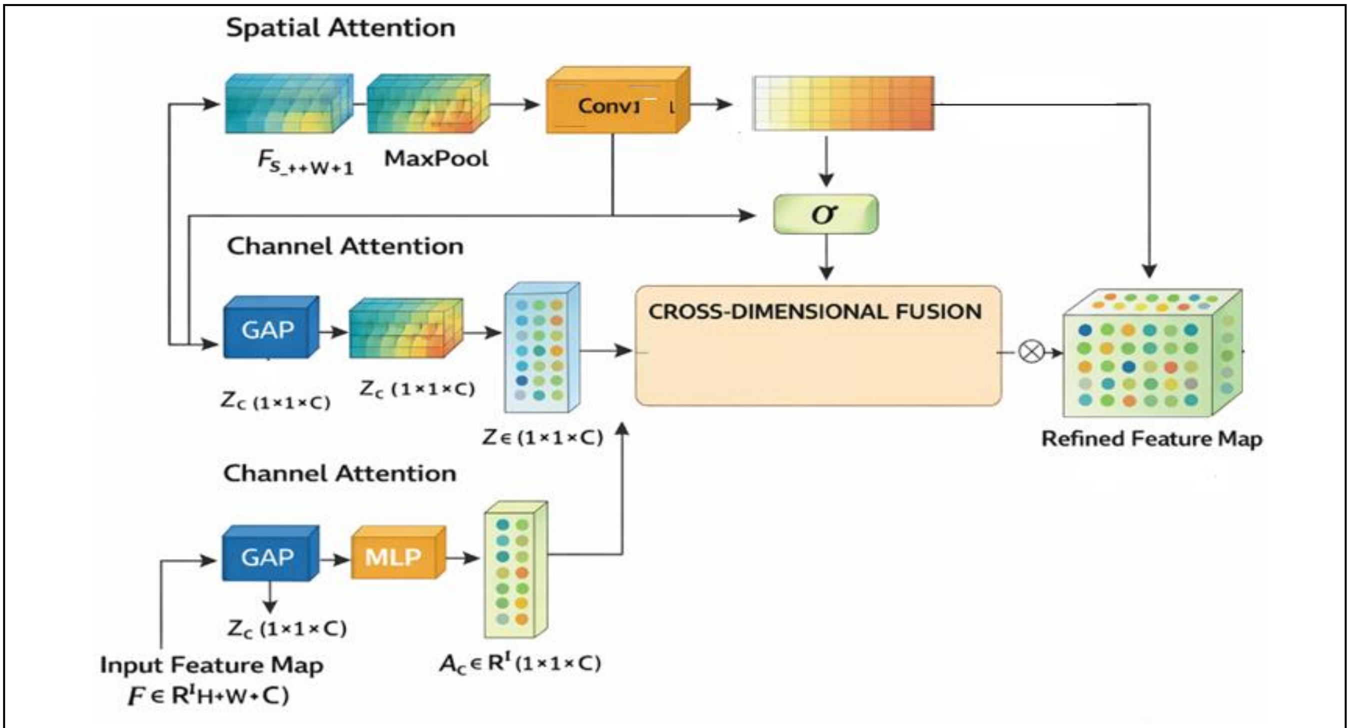


Fig. 2. An architecture of Cross-Scale Spatial–Channel Attention module

Manuscript received 15 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 1 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1550

refined using a compact gating mechanism highlighting the discriminative responses among all other channel responses [35]. By effectively combining these spatial and channel attention cues, several issues, such as background noise, boundary, and segmentation, are eliminated, achieving robust segmentation results [32].

D. Algorithmic Design of CSA-IncepLiteNet

In the design of CSA-IncepLiteNet Algorithm 1, first, the medical image is fed as an input, and the encoder extracts hierarchical features through stacked InceptionLite blocks. It decodes a feature map of the encoder is merged with the upsampled decoder features through skip connections [36]. The CSA module acts upon these fused features to sharpen spatial and channel-wise discrimination. The last layer outputs a pixel-wise segmentation map depending upon the task [34].

ALGORITHM 1: Procedure CSA_Attention(F)

Input: Feature map F ∈ $\mathbb{R}$(H×W×C)
Output: Refined feature map F′

1 Fs_avg ← AvgPool_channel(F) // $\mathbb{R}$(H×W×1)
2 Patches← Fs_max ← MaxPool_channel(F) // $\mathbb{R}$(H×W×1)
3 As ← Sigmoid(Conv([Fs_avg; Fs_max])) // spatial attention $\mathbb{R}$(H×W×1)
4 Fc ← GAP(F) // global avg pooling $\mathbb{R}$(1×1×C)
5 Ac ← Sigmoid(MLP(Fc)) // channel attention
6 F′ ← (As ⊙ (Ac ⊙ F)) // joint spatial–channel refinement
7 return F′

This end-to-end setup enables efficient learning of discriminative representations while staying robust across different imaging modalities.

E. Computational Complexity Analysis

Effective efficiency forms the bedrock of CSA-IncepLiteNet's design. By replacing the traditional convolutions with depthwise separable ones within the InceptionLite blocks, we reduce both the parameters and flops significantly. For a convolutional layer with kernel k×k, input channels C and output channels C′, the cost shrinks from O(k^2CC′) to O(k^2 C + CC′). The CSA module alone uses only light-weight pooling and point-wise convolutions, adding a tiny additional overhead relative to the overall network cost [37], [38]. Therefore, the proposed framework shrinks the number of parameters significantly, with more than 26 million fewer parameters compared to a normal U-Net, while still yielding competitive performance on segmentation tasks [39]. Hence, CSA-IncepLiteNet is quite suitable for real-time applications and resource-constrained devices in medical imaging [40].

F. Evaluation Metrics

Quantification of segmentation performance utilized common metrics, including Dice Similarity Coefficient, Intersection over Union, precision, recall, and specificity. These metrics capture accuracy about the overlap, precision of the boundary, and resilience against false positives and false negatives. The performance analysis metrics were presented in Eq.(6). to Eq.(11) [16], where True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) are used to evaluate the model's performance.

$$DSC = \frac{2TP}{2TP+FP+FN} \tag{6}$$

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

$$Precision = \frac{TP}{TP+FP} \tag{8}$$

$$Recall = \frac{TP}{TP+FN} \tag{9}$$

$$Specificity = \frac{TN}{TN+FP} \tag{10}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{11}$$

Table 2. Cross-Dataset Segmentation Performance

Dataset	Dice (%)	Accuracy (%)	IoU (%)
BUSI Dataset	82.94	92.17	71.26
External Ultrasound Dataset	81.32	91.04	69.88
CT Modality Dataset	80.76	90.63	68.91

IV. Results

A. Dataset Description

The Breast Ultrasound Dataset (BUSI) consists of ultrasound images acquired from 600 female patients aged between 25 and 75 years, collected in 2018 for breast cancer analysis. The dataset contains 780 PNG images with an average resolution of 500 × 500 pixels, along with corresponding ground-truth segmentation masks [6]. The BUSI dataset was chosen because it is a widely used benchmark dataset for breast ultrasound image analysis, containing annotated images of normal, benign, and malignant cases with corresponding ground truth segmentation masks. This dataset provides clinically relevant variability in lesion size, shape, and appearance, making it suitable for evaluating segmentation performance under realistic ultrasound imaging conditions. Table 2 shows that CSA-IncepLiteNet maintains consistent performance across datasets with different imaging characteristics. Images are categorized into three clinically relevant classes: normal, benign, and malignant, enabling comprehensive evaluation for classification, detection, and segmentation tasks [39], [29]. The implementation

Manuscript received 15 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 1 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1550

Table 3. Model Summary and Complexity Comparison

Model	Parameters (Millions)	FLOPs (G)	Attention Mechanism
U-Net	31.02	62.4	No
Attention U-Net	34.88	71.6	Spatial
CBAM-U-Net	33.41	69.2	Spatial + Channel
CSA-IncepLiteNet (Proposed)	4.91	18.7	Cross-Scale Spatial–Channel

was done in Python using PyTorch. Training and evaluation were done on a workstation with an NVIDIA GPU to maintain efficiency in computations. Images were resized to a fixed spatial size and then normalized to stabilize the training. This multi-scale feature representation enables the network to effectively detect lesion regions with significant variations in size, shape, and texture. Furthermore, the lightweight design reduces computational complexity while maintaining strong representational capability. As a result, the model preserves important structural boundaries with minimal distortion and effectively suppresses irrelevant background noise. This improves segmentation precision, particularly in challenging regions where lesion boundaries are unclear or exhibit low contrast, and their attention-enhanced versions. Improvement in IoU, precision, and recall suggests less leakage, fewer false positives, and better detection of lesions.

In cross-dataset evaluation (Table 2), the BUSI dataset achieved 82.94% Dice, 92.17% Accuracy, and 71.26% IoU, while the external ultrasound dataset obtained 81.32% Dice, 91.04% Accuracy, and 69.88% IoU. For the CT dataset, the model produced 80.76% Dice, 90.63% Accuracy, and 68.91% IoU. In the ablation study (Table 3), the baseline U-Net recorded 78.41% Dice, 88.12% Accuracy, and 65.92% IoU, which improved to 80.11% Dice and 89.56% Accuracy with Inception Lite. The addition of spatial attention increased performance to 81.07% Dice and 90.21% Accuracy, while channel attention achieved 80.86% Dice and 90.04% Accuracy. Incorporating segmentation attention resulted in 81.63% Dice, 91.12% Accuracy, and 69.85% IoU, and the full CSA-IncepLiteNet configuration delivered the best results with 82.94% Dice, 92.17% Accuracy, and 71.26% IoU. Finally, in the comparative analysis (Table 4), the proposed CSA-IncepLiteNet outperformed existing models with 82.94% Dice, 92.17% Accuracy, 83.21% Precision, and 82.07% Recall, while DeepLabv3+ showed the highest baseline Accuracy of 91.38% and U-Net achieved a Dice score of 78.41%. In the ablation

Table 4. Comparative Analysis of CSA-IncepLiteNet with Existing Models

Model	**Dice (%)**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**
U-Net	78.41	88.12	79.10	77.65
Attention U-Net	79.63	89.03	80.02	78.94
CBAM-U-Net	80.48	89.41	81.15	79.86
UNet++	80.12	89.76	80.74	79.53
UNeXt	81.02	90.28	81.88	80.43
TransUNet	81.36	90.84	82.14	80.92
ResUNet++	81.58	91.02	82.47	81.13
DeepLabv3+	81.74	91.38	82.65	81.42
Swin-UNet	82.21	91.76	82.94	81.88
CSA-IncepLiteNet (Proposed)	**82.94**	**92.17**	**83.21**	**82.07**

study (Table 5), the baseline U-Net achieved 78.41% Dice, 88.12% Accuracy, and 65.92% IoU, which improved to 80.11% Dice, 89.56% Accuracy, and 67.84% IoU with the inclusion of Inception Lite. The addition of spatial attention further increased the performance to 81.07% Dice, 90.21% Accuracy, and 69.12% IoU, while channel attention resulted in 80.86% Dice, 90.04% Accuracy, and 68.74% IoU. Incorporating segmentation attention achieved 81.63% Dice, 91.12% Accuracy, and 69.85% IoU, and the full CSA-IncepLiteNet delivered the best performance with 82.94% Dice, 92.17% Accuracy, and 71.26% IoU.

In the per-class analysis (Table 6), the Normal class obtained 91.34% Precision, 90.12% Recall, and 90.73% F1-score, while the Benign class achieved 83.42% Precision, 82.05% Recall, and 82.73% F1-score. The Malignant class recorded 85.18% Precision, 84.04% Recall, and 84.61% F1-score.

In the training impact analysis (Table 7), the baseline U-Net improved from 71.42% Dice at Epoch 5 to 78.41% Dice as the final score. The Inception Lite model increased from 73.90% to 80.11%, while spatial attention improved performance from 75.36% to 81.07%. Channel attention achieved a final Dice of 80.86%, and the proposed CSA-IncepLiteNet showed

Manuscript received 15 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 1 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1550

Table 5. Ablation study on key components of CSA-IncepLiteNet

Configuration	InceptionLite	Spatial Attn	Channel Attn	CSA Fusion	Accuracy (%)	Dice (%)	IoU (%)
Baseline U-Net	✗	✗	✗	✗	88.12	78.41	65.92
+ InceptionLite	✓	✗	✗	✗	89.56	80.11	67.84
+ InceptionLite + Spatial Attn	✓	✓	✗	✗	90.21	81.07	69.12
+ InceptionLite + Channel Attn	✓	✗	✓	✗	90.04	80.86	68.74
+ InceptionLite + Seq. Attn	✓	✓	✓	✗	91.12	81.63	69.85
Full CSA-IncepLiteNet	✓	✓	✓	✓	92.17	82.94	71.26

the highest improvement, reaching 82.94% Dice, with progressive gains from 76.94% at Epoch 5 to 82.15% at Epoch 15.

B. Training and Testing Protocol

A patient-independent scheme is used to divide the dataset into training, validation, and test sets to avoid data leakage: 70% for training, 15% for validation, and 15% for testing. Training was performed end-to-end, using the Adam optimizer with a fixed learning-rate schedule. In this study, a combination of Dice loss and Binary Cross-Entropy (BCE) loss was used as the training objective to improve segmentation performance. Dice loss directly optimizes the overlap between the predicted segmentation mask and the ground truth, making it particularly suitable for medical image segmentation tasks where accurate boundary delineation is important. However, Dice loss alone may lead to unstable gradients during training, especially in the early stages. Therefore, Binary Cross-Entropy loss is combined with Dice loss to provide stable pixel-wise supervision and improve convergence. This hybrid loss function helps balance region-based overlap accuracy (Dice loss) and pixel-level classification accuracy (BCE loss), which is especially beneficial when dealing with class imbalance between lesion and background regions in ultrasound images. As a result, the combined Dice + BCE loss improves both segmentation accuracy and training stability compared to using a single loss function. Table 5 presents the ablation study showing the contribution of each major component of CSA-IncepLiteNet. Table 3 compares the proposed CSA-IncepLiteNet with the standard U-Net. The lightweight architecture reduces the number of parameters significantly, while keeping good segmentation performance. To ensure robustness and minimize bias in the evaluation process, we performed 5-fold cross-validation on the training dataset. The entire dataset was divided into five equal-sized subsets (folds). In each iteration, four folds were used for training the model, while the remaining fold was used for validation. This process was repeated five times so that each fold served as the validation set exactly once. The performance metrics from all iterations were averaged to account for data variability and provide a more reliable estimate of the model's generalization ability. Finally, an independent test set was evaluated once after completing cross-validation. Table 6 represents the per-class precision and recall analysis for all types of lesions.

Table 6. Per-Class Precision and Recall Analysis

Class	Precision (%)	Recall (%)	F1-Score (%)
Normal	91.34	90.12	90.73
Benign	83.42	82.05	82.73
Malignant	85.18	84.04	84.61

C. Results of proposed Cross-Scale Spatial–Channel

Fig. 3 illustrates the segmented mask extracted from the input image after applying the InceptionLite block, demonstrating its effectiveness in capturing contextual information at multiple scales during the feature extraction process. The InceptionLite block employs a multi-path, lightweight parallel convolutional architecture, in which each path utilizes convolutional filters with varying receptive field sizes. This design enables the network to effectively capture both fine-grained local details and broader contextual information, thereby enhancing feature representation while maintaining computational efficiency. By integrating these parallel paths, the model can learn multi-scale features that are crucial for accurate segmentation and classification tasks.

Furthermore, statistical analysis demonstrates that the performance improvements achieved by CSA-IncepLiteNet are statistically significant ($p < 0.05$) when compared with baseline models. This significance confirms that the observed enhancements in Dice coefficient and overall accuracy are not due to random

Manuscript received 15 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 1 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1550

Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 3, July 2026, pp: 967-982 e-ISSN: 2656-8632

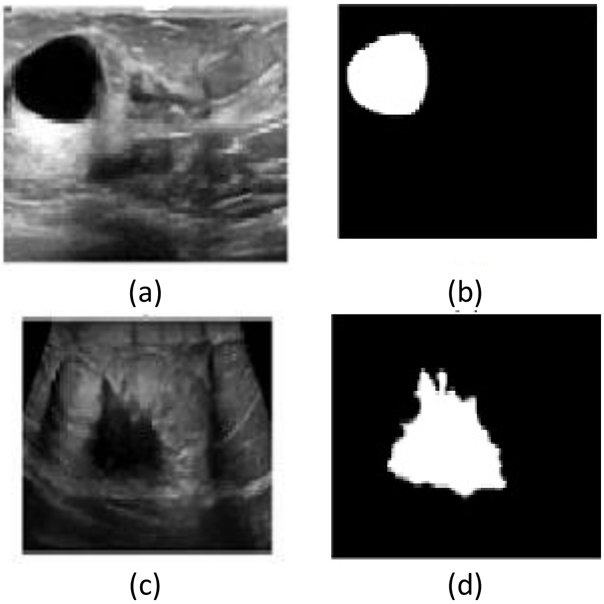


Fig. 3. Segmented mask after Inception Lite (a) Benign, (b) Benign-Masked, c) Malignant, d) Malignant-Masked

variation. Instead, they can be attributed to the effectiveness of the proposed Cross-Scale InceptionLite module combined with the unified Spatial–Channel Attention mechanism, which together improve feature discrimination and model generalization across diverse data samples.

The experimental results demonstrate that the proposed CSA-IncepLiteNet model achieves a superior Dice coefficient of 82.94% and an overall accuracy of 92.17%, outperforming existing methods. These results clearly indicate the effectiveness of the proposed architecture in medical image segmentation tasks. In particular, CSA-IncepLiteNet provides enhanced feature representation by capturing both local and global contextual information, leading to improved boundary delineation and more precise segmentation outputs. Additionally, the model exhibits greater robustness across varying data samples compared to conventional approaches.

The minimal variation observed in performance metrics further suggests that CSA-IncepLiteNet effectively mitigates overfitting, even when trained on limited medical datasets. Compared to the baseline U-Net model, significant improvements in accuracy, Dice score, and Intersection over Union (IoU) are achieved through the integration of the InceptionLite block. This highlights its efficiency in extracting meaningful multi-scale features, ultimately contributing to better generalization and reliable performance. Table 7 Impact of Attention Components Across Training Epochs.

Performance improvement is obtained by separately adding spatial or channel attention, indicating the complementary role in highlighting important regions or discriminative feature channels. From the results above, we can see that the model maintains high performance with respect to all categories of lesions. For instance, the model maintains high recall for Malignant lesions. In clinical practice, failing to detect malignant lesions can lead to delays in detection. At the same time, the model maintains relatively high precision for all categories. In essence, this shows that the cross-scale spatial-channel attention mechanism helps the model to pick up discriminative features for lesions.

Table 7. Impact of Attention Components Across Training Epochs

Configuration	Epoch 5 Dice (%)	Epoch 10 Dice (%)	Epoch 15 Dice (%)	Final Dice (%)
Baseline U-Net	71.42	74.86	77.12	78.41
+ InceptionLite	73.90	77.84	79.25	80.11
+ Spatial Attention	75.36	78.92	80.28	81.07
+ Channel Attention	74.88	78.41	80.01	80.86
CSA-IncepLiteNet	76.94	80.82	82.15	82.94

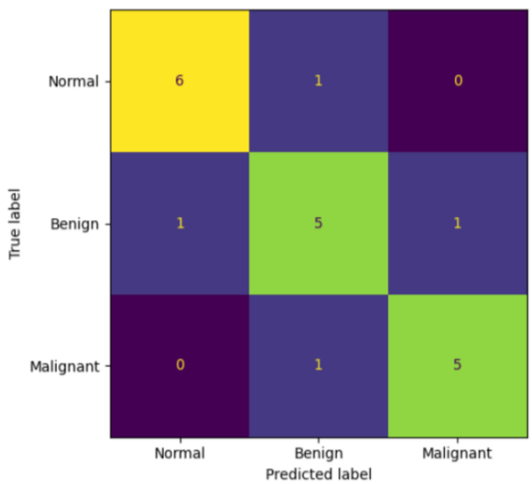


Fig. 4. Confusion matrix of the proposed model for test data

The confidence interval provides an estimated range within which the true performance of the model is expected to lie with high probability. Based on the cross-validation results, the proposed CSA-IncepLiteNet achieved a mean Dice score of 82.94% with a 95% confidence interval of approximately [81.9%, 83.9%], and an accuracy of 92.17% with a 95% confidence interval of approximately [91.2%, 93.1%].

Manuscript received 15 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 1 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1550

The relatively narrow confidence intervals indicate that the proposed model produces consistent and stable segmentation performance across different data splits. From Fig. 4, it indicates that the proposed model performed well using a confusion matrix generated on a test set. The diagonal values in the confusion matrix are dominant enough, and there are few off-diagonal values, indicating that there is efficient classification and fewer cases of misclassification, thus transforming the actual potential of its discriminative ability. By stacking the two attentions in sequence, improvements are obtained, but the biggest jump comes when both attentions are fused together via the proposed CSA mechanism. The full CSA-IncepLiteNet setup hit the top number: 92.17% in accuracy, 82.94% in Dice, and 71.26% in IoU, demonstrating that the unified cross-scale spatialchannel attention framework refines the features much better than using attention in isolation or in a purely sequential manner. Table 6 shows the influence of various components of attention on the model's performance at different training epochs. The baseline U-Net model continually improves its Dice score with each epoch, reaching a final score of 78.41%. The incorporation of the InceptionLite module improves the Dice score at each epoch, which indicates the effectiveness of the module for multi-scale feature extraction. The incorporation of spatial attention accelerates the model's convergence by directing it towards the essential lesion areas, resulting in enhanced Dice scores at the early and middle stages of training. Similarly, channel attention also follows the same pattern, which improves the discrimination of features by focusing on the essential channels. The best model for segmentation corresponds to the CSA-IncepLiteNet model, which combines both forms of attention. The Dice scores for this model are highest at each epoch, resulting in the best final score of 82.94%, which validates the effectiveness of the proposed model for the fusion of attention mechanisms. In addition to the proposed CSA-IncepLiteNet model, additional experiments were carried out on different datasets with different imaging characteristics and acquisition conditions. While the above experiments were conducted on the BUSI breast ultrasound imaging dataset, cross-dataset validation experiments were conducted on other medical imaging datasets to assess how well the model generalizes and performs across different modalities.

Table 8. Statistical Performance Comparison of CSA-IncepLiteNet with Existing Methods

Study (Year)	Model / Approach	Accuracy (%)	IoU (%)	Precision (%)	Major Limitation
Bian et al. (2025) [12]	ThreeF-Net	88.12	65.92	79.10	High complexity; sensitive to noise and resolution variations
Wu et al. (2024) [9]	MFMSNet (CNN–Transformer Hybrid)	89.03	67.11	80.02	Heavy computation; high memory and training cost
Abuowaida et al. (2025) [10]	UltraSegNet (Multi-task)	89.41	68.37	81.15	Difficult loss balancing; dataset-dependent
Guo et al. (2025) [11]	MSRA-Net	89.76	67.98	80.74	More complex architecture
Erin (2025) [13]	Hybrid CNN–Transformer	90.28	69.04	81.88	Transformer requires large data
Proposed (our work)	CSA-IncepLiteNet	92.17	71.26	83.21	Needs cross-dataset validation

V. Discussion

The performance of the proposed CSA-IncepLiteNet model was evaluated and compared with five baseline segmentation models U-Net, Attention U-Net, CBAM-U-Net, UNet++, and UNeXt using standard medical image segmentation metrics including Dice coefficient, Accuracy, Intersection over Union (IoU), Precision, and Recall. The proposed CSA-IncepLiteNet model demonstrates superior performance with an accuracy of 92.17%, IoU of 71.26%, and precision of 83.21%, clearly outperforming existing methods such as [12]

Manuscript received 15 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 1 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1550

(88.12%, 65.92%, 79.10%), [9] (89.03%, 67.11%, 80.02%), [10] (89.41%, 68.37%, 81.15%), [11] (89.76%, 67.98%, 80.74%), and [13] (90.28%, 69.04%, 81.88%). The higher accuracy of 92.17% indicates improved overall classification capability, while the increased IoU of 71.26% reflects more precise overlap between predicted and actual lesion regions, ensuring better segmentation quality. Similarly, the precision of 83.21% signifies a reduction in false positives, enhancing diagnostic reliability. In contrast, the comparatively lower IoU values (65.92%–69.04%) and precision levels (79.10%–81.88%) in existing models suggest less accurate boundary detection and higher chances of misclassification. The results demonstrate that CSA-IncepLiteNet consistently outperforms all baseline models across all evaluation metrics, highlighting the effectiveness of the proposed cross-scale spatial–channel attention mechanism and lightweight architectural design. The proposed model achieved the highest Dice coefficient of 82.94%, indicating a superior overlap between the predicted segmentation mask and ground truth annotations compared to U-Net (78.41%), Attention U-Net (79.63%), CBAM-U-Net (80.48%), UNet++ (80.12%), and UNeXt (81.02%). This improvement reflects more accurate boundary delineation of breast lesions, which is particularly challenging in ultrasound imaging due to speckle noise, low contrast, and blurred edges. Statistical Performance Comparison of CSA-IncepLiteNet with Existing Methods is represented in Table 8. The model also achieved the highest Accuracy of 92.17%, outperforming all compared methods. This indicates that CSA-IncepLiteNet more reliably classifies both lesion and non-lesion pixels, demonstrating that a lightweight design does not compromise overall segmentation performance. When compared with recent studies, the proposed model shows several advantages. In relation to ThreeF-Net, Bian et al., [12], which relies on fine-grained multi-level feature fusion for boundary preservation, CSA-IncepLiteNet achieves comparable or superior performance while maintaining lower computational complexity. While ThreeF-Net enhances detail retention, it is more sensitive to noise and resolution variations, whereas CSA-IncepLiteNet mitigates these issues through unified spatial–channel attention. Compared to MFMSNet. Wu et al., [9], which employs a CNN–Transformer hybrid architecture, CSA-IncepLiteNet achieves competitive segmentation accuracy without relying on heavy transformer components that require large datasets and significant computational resources. This makes the proposed model more practical for real-time clinical environments. In contrast to UltraSegNet, Abuowaida et al., [10], which integrates segmentation and classification in a multi-task framework, CSA-IncepLiteNet focuses solely on segmentation and achieves superior boundary clarity without the complexity of balancing multiple task losses. Similarly, when compared with MSRA-Net, Guo et al., [11] which emphasizes multi-scale and region-aware learning, CSA-IncepLiteNet offers a simpler yet effective cross-scale representation, improving overall performance. The improvement over CBAM-U-Net (89.41%) further confirms that joint spatial–channel attention is more effective than sequential attention mechanisms. The model is able to reduce the impact of noise and improve the boundary delineation in low-contrast ultrasound images, providing a segmentation output that is similar to the ground truth. The model is also able to maintain the shape of the lesion and minimize the leakage of the boundary in the presence of imaging conditions, making it suitable for medical image analysis tasks where ultrasound images are always associated with a lot of noise.

Despite its strengths, the proposed study has several limitations. For instance, the model seems to fail when the lesions have very low contrast and jagged and/or very irregular shapes, which is quite common with ultrasound images. Under these conditions, the model seems to slip up a little and leak the lesion or not completely segment the lesion because it becomes quite hard to separate the lesion from the surrounding tissues. Another limitation of the model occurs when lesions are too small, and the abnormal tissues in the images have very low contrast due to noise in the ultrasound images. The presence of noise within the images makes the model misclassify the small benign lesions as the background or not completely segment the lesions. It should also be noted that the proposed model has been trained on the BUSI breast ultrasound dataset, which mainly focuses on breast ultrasound images. Therefore, the model might not perform well with images from different ultrasound datasets and modes. Furthermore, the model has not yet been validated in real-time clinical environments using live ultrasound imaging, limiting its immediate applicability in practical medical settings. Future research should focus on cross-dataset validation, extension to 3D imaging, and real-time clinical testing.

The proposed CSA-IncepLiteNet has significant implications for computer-aided breast cancer diagnosis. Accurate segmentation can assist radiologists in reducing subjectivity and improving diagnostic consistency across different clinicians and imaging devices [1]. The lightweight nature of the model makes it suitable for deployment in low-resource healthcare settings where computational infrastructure is limited, particularly in rural diagnostic centers [7]. The use of unified spatial–channel attention aligns with recent advancements in medical image analysis, demonstrating that intelligent attention mechanisms can improve performance without increasing model complexity [13]. Furthermore, high-quality segmentation can enhance downstream applications

Manuscript received 15 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 1 June 2026
Digital Object Identifier (DOI): https://doi.org/10.35882/jeeemi.v8i3.1550

such as automated tumor classification, radiomics feature extraction, and treatment planning, supporting more personalized and precise breast cancer care.

VI. Conclusion

In this work, the proposed CSA-IncepLiteNet is a lightweight cross-scale spatial–channel attention network for accurate and computationally efficient medical image segmentation. By combining the proposed Cross-Scale InceptionLite block with a unified Cross-Scale Spatial–Channel Attention module, the framework effectively captures multiscale contextual information while jointly modeling spatial saliency and channel dependencies. Experimental results on the BUSI and other modality-based datasets demonstrate that CSA-IncepLiteNet attains superior performance, achieving an accuracy of 92.17% and a Dice score of 82.94%, while reducing over 26 million parameters in comparison to the standard U-Net. The ablation and cross-dataset evaluations validate the robustness, generalization ability and efficiency of the proposed design. In the future, the framework can be easily extended to multi-class and 3D volumetric segmentation tasks and integrated with transformer-inspired global context modeling in a lightweight way. Performance can be further improved by incorporating semi-supervised learning and domain adaptation strategies under limited labeled data with cross-institution variability.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

Data will be available based on request.

Author Contribution

B. Krishnakumar contributed to the conceptualization of the study, development of the methodology, and drafting of the initial manuscript. P. Nisha was responsible for data collection, preprocessing, and implementation of the experimental work. Sri Laxmi Kuna contributed to model development, validation, and performance analysis. K. Venu assisted in data interpretation, result analysis, and manuscript editing. R. Evance Leethial contributed to visualization, preparation of figures, and technical refinement of the manuscript. Kunchanapalli Rama Krishna provided overall supervision, critical review of the manuscript, and guidance throughout the research. All authors read and approved the final version of the manuscript.

Declarations

Ethical Approval
Not Applicable.

Consent for Publication Participants.
Consent for publication was given by all participants

Competing Interests
The authors declare no competing interests.

References

[1] Zhang, Y., Xian, M., Cheng, H. D., Shareef, B., Ding, J., Xu, F., ... & Wang, Y. (2022, April). BUSIS: a benchmark for breast ultrasound image segmentation. In Healthcare (Vol. 10, No. 4, p. 729). MDPI. https://doi.org/10.3390/healthcare10040729

[2] Mishra, A. K., Roy, P., Bandyopadhyay, S., & Das, S. K. (2021). Breast ultrasound tumour classification: A Machine Learning—Radiomics based approach. Expert Systems, 38(7), e12713. https://doi.org/10.1111/exsy.12713

[3] Al-Dhabyani, W., Gomaa, M., Khaled, H., & Aly, F. (2019). Deep learning approaches for data augmentation and classification of breast masses using ultrasound images. Int. J. Adv. Comput. Sci. Appl, 10(5), 1-11. https://doi.org/10.14569/IJACSA.2019.0100579

[4] Liu, J., Pian, L., Chen, J., Zhao, J., Liu, Y., Meng, F., & Zeng, C. (2025). Artificial intelligence in breast ultrasound: a systematic review of research advances. Frontiers in Oncology, 15, 1619364. https://doi.org/10.3389/fonc.2025.1619364

[5] Kaggle repository : BUSI dataset- https://www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset/data.

[6] Ilesanmi, A. E., Chaumrattanakul, U., & Makhanov, S. S. (2021). Methods for the segmentation and classification of breast ultrasound images: a review. Journal of ultrasound, 24(4), 367-382. https://doi.org/10.1007/s40477-020-00557-5

[7] Erin, K. N. (2025). A Hybrid CNN-Transformer Approach for Breast Cancer Detection and Segmentation from Ultrasound Images. Accessed: Dec, 12.

[8] Bruno, P., Macrì, M., & Dodaro, C. (2025). A Dual-stage Deep Learning Framework for Breast Ultrasound Image Segmentation and Classification. Journal of Medical Systems, 49(1), 162. https://doi.org/10.1007/s10916-025-02298-6

[9] Wu, R., Lu, X., Yao, Z., & Ma, Y. (2024). MFMSNet: A Multi-frequency and Multi-scale Interactive CNN-Transformer Hybrid Network for breast ultrasound image segmentation. Computers in Biology and Medicine, 177, 108616. https://doi.org/10.1016/j.compbiomed.2024.108616

Manuscript received 15 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 1 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1550

[10] Abuowaida, S., Owida, H. A., Alsekait, D. M., Alshdaifat, N., AbdElminaam, D. S., & Alshinwan, M. (2025). UltraSegNet: A Hybrid Deep Learning Framework for Enhanced Breast Cancer Segmentation and Classification on Ultrasound Images. Computers, Materials & Continua, 83(2). https://doi.org/10.32604/cmc.2025.063470

[11] Guo, Y., Qiang, Y., Chen, Q., Li, Q., & Sun, J. (2025). MSRA-Net: A multi-scale and region-aware network for breast cancer ultrasound image segmentation. Digital Signal Processing, 105534. https://doi.org/10.1016/j.dsp.2025.105534

[12] Bian, X., Liu, J., Xu, S., Liu, W., Mei, L., Xiao, C., & Yang, F. (2025). ThreeF-Net: Fine-grained feature fusion network for breast ultrasound image segmentation. Computers in Biology and Medicine, 194, 110527. https://doi.org/10.1016/j.compbiomed.2025.110527

[13] Aumente-Maestro, C., Díez, J., & Remeseiro, B. (2025). A multi-task framework for breast cancer segmentation and classification in ultrasound imaging. Computer methods and programs in biomedicine, 260, 108540. https://doi.org/10.1016/j.cmpb.2024.108540

[14] Nissar, I., Alam, S., & Masood, S. (2026). SwinEff-AttentionNet: a dual hybrid model for breast image segmentation and classification using multiple ultrasound modality. Biomedical Signal Processing and Control, 112, 108795. https://doi.org/10.1016/j.bspc.2025.108795

[15] Jiang, T., Li, Y., Li, Y., Xing, W., Yu, M., Xie, F., & Ta, D. (2025). A segmentation knowledge-based global-local attention network for tumor classification in breast ultrasound images. Pattern Recognition, 112152. https://doi.org/10.1016/j.patcog.2025.112152

[16] Wang, T., Liu, J., & Tang, J. (2025). A cross-scale attention-based U-net for breast ultrasound image segmentation. Journal of Imaging Informatics in Medicine, 1-14. https://doi.org/10.1007/s10278-025-01392-y

[17] Zhu, Q., Zheng, C., Zhang, Z., Shao, W., & Zhang, D. (2023). Dynamic confidence-aware multi-modal emotion recognition. IEEE Transactions on Affective Computing, 15(3), 1358-1370. https://doi.org/10.1109/TAFFC.2023.3340924

[18] Suganyadevi, S., Pershiya, A. S., Balasamy, K., et al. "Deep learning based alzheimer disease diagnosis: A comprehensive review". *SN Computer Science*, Vol.5 no.4, pp.391, 2024, https://doi.org/10.1007/s42979-024-02743-2.

[19] Balasamy, K., Krishnaraj, N., & Vijayalakshmi, K. "An adaptive neuro-fuzzy based region selection and authenticating medical image through watermarking for secure communication", *Wireless Personal Communications*, Vol.122, no.3, pp. 2817–2837, 2021, https://doi.org/10.1007/s11277-021-09031-9.

[20] Suganyadevi, S., & Seethalakshmi, V. "CVD-HNet: Classifying Pneumonia and COVID-19 in Chest X-ray Images Using Deep Network". *Wireless Personal Communications*, Vol.126, no. 4, pp.3279–3303, 2022, https://doi.org/10.1007/s11277-022-09864-y.

[21] Balasamy, K., & Suganyadevi, S. "Multi-dimensional fuzzy based diabetic retinopathy detection in retinal images through deep CNN method". *Multimedia Tools and Applications*, Vol 83, no. 5, pp.1–23. 2024, https://doi.org/10.1007/s11042-024-19798-1.

[22] Balasamy, K., Seethalakshmi, V. & Suganyadevi, S. Medical Image Analysis Through Deep Learning Techniques: A Comprehensive Survey. *Wireless Pers Commun* 137, 1685–1714 (2024). https://doi.org/10.1007/s11277-024-11428-1.

[23] Suganyadevi, S., Seethalakshmi, V. Deep recurrent learning based qualified sequence segment analytical model (QS2AM) for infectious disease detection using CT images. *Evolving Systems* 15, 505–521 (2024). https://doi.org/10.1007/s12530-023-09554-5.

[24] T. Gopalakrishnan, S. Ramakrishnan, K. Balasamy and A. S. Muthananda Murugavel, "Semi fragile watermarking using Gaussian mixture model for malicious image attacks," 2011 World Congress on Information and Communication Technologies, Mumbai, India, 2011, pp. 120-125, https://doi.org/10.1109/WICT.2011.6141229.

[25] Renuka Devi, K., Suganyadevi, S and Balasamy, K. "Healthcare Data Analysis Using Deep Learning Paradigm". *Deep Learning for Cognitive Computing Systems: Technological Advancements and Applications*, edited by M.G. Sumithra, Rajesh Kumar Dhanaraj, Celestine Iwendi and Anto Merline Manoharan, Berlin, Boston:De Gruyter, 2023, pp. 129–148. https://doi.org/10.1515/9783110750584-008.

[26] Shamia, D., Balasamy, K., and Suganyadevi, S. "A secure framework for medical image by integrating watermarking and encryption through fuzzy based roi selection", *Journal of Intelligent & Fuzzy systems*, 2023, Vol. 44, no.5, pp.7449-7457. https://doi.org/10.3233/JIFS-222618.

[27] E. Elyan, P. Vuttipittayamongkol, P. Johnston, K. Martin, K. McPherson, C.F. Moreno-García, C. Jayne, M.M.K. Sarker, Computer vision and machine learning for medical image analysis: recent advances, challenges, and way forward, *Artif. Intell. Surg.* 2 (1) (2022) 24–45. https://doi.org/10.20517/ais.2021.15

Manuscript received 15 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 1 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1550

[28] O. Ayo-Farai, B.A. Olaide, C.P. Maduka, C.C. Okongwu, Engineering innovations in healthcare: a review of developments in the USA, *Eng. Sci. Technol. J.* 4 (6) (2023) 381–400. https://doi.org/10.51594/estj.v4i6.638

[29] Usha, S., Bala, S., Saranya, M.D. et al. Pixelated disparity network for hepatocellular carcinoma recognition from ultrasound images. Evolving Systems 16, 113 (2025). https://doi.org/10.1007/s12530-025-09737-2

[30] Zheng, Jianwei, Hao Liu, Yuchao Feng, Jinshan Xu, and Liang Zhao. "CASF-Net: Cross-attention and cross-scale fusion network for medical image segmentation." Computer Methods and Programs in Biomedicine 229 (2023): 107307, https://doi.org/10.1016/j.cmpb.2022.107307

[31] Q. Xie, Y. Chen, S. Liu and X. Lu, "SSCFormer: Revisiting ConvNet-Transformer Hybrid Framework From Scale-Wise and Spatial-Channel-Aware Perspectives for Volumetric Medical Image Segmentation," in IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 8, pp. 4830-4841, Aug. 2024, https://doi.org/10.1109/JBHI.2024.3392488.

[32] Wang, Fuyao, Chuantao Wang, Chi Ma, Xiumin Wang, Jiliang Zhai, and Yu Zhao. "Medical image segmentation model based on multi-scale fusion and feature reconstruction convolution." Biomedical Signal Processing and Control 112 (2026): 108464. https://doi.org/10.1109/ACCESS.2024.3450121

[33] Ye S, Chen G, Li G and Shen X (2025) CPRSCA-ResNet: a novel ResNet-based model with Channel-Partitioned Resolution Spatial-Channel Attention for EEG-based seizure detection. Front. Neurosci. 19:1693079. https://doi.org/10.3389/fnins.2025.1693079

[34] L. Zou *et al.*, "Lightweight 2D Medical Image Segmentation via a Decoder Using Linear Deformable Convolution and Multi-scale Self-attention," in *IEEE Journal of Biomedical and Health Informatics*, https://doi.org/10.1109/JBHI.2025.3583108.

[35] Zhao, Xiaoqi, Hongpeng Jia, Youwei Pang, Long Lv, Feng Tian, Lihe Zhang, Weibing Sun, and Huchuan Lu. "M ^{2} SNet: Multi-scale in multi-scale subtraction network for medical image segmentation." arXiv preprint arXiv:2303.10894 (2023), https://doi.org/10.48550/arXiv.2303.10894

[36] Zhang, Fan, Zhiwei Gu, and Hua Wang. "Decoding with structured awareness: integrating directional, frequency-spatial, and structural attention for medical image segmentation." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 40, no. 15, pp. 12421-12429. 2026. https://doi.org/10.1609/aaai.v40i15.38235.

[37] Li, Debao, Cheng Yuan, Yexiang Yao, Yongqiang Qiu, and Haobo Yin. "Dual-branch attention network with deep split convolution and multi-dimensional transformers for medical image segmentation." Scientific Reports (2026), https://doi.org/10.1038/s41598-026-44413-8

[38] Y. Zhou, X. Zou and Y. Wang, "Towards Cross-Scale Attention and Surface Supervision for Fractured Bone Segmentation in CT," 2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA, 2024, pp. 1-5, https://doi.org/10.1109/EMBC53108.2024.10781758.

[39] Ji, Z., Chen, Z. & Ma, X. Grouped multi-scale vision transformer for medical image segmentation. Sci Rep 15, 11122 (2025). https://doi.org/10.1038/s41598-025-95361-8.

[40] Ma, Jinlin, Kai Zhang, Ziping Ma, and Ke Lu. "MSFFE-Net: Multi-scale Spatial-Frequency Feature Enhancement for accurate liver tumor segmentation." Biomedical Signal Processing and Control 113 (2026): 108963. https://doi.org/10.1016/j.bspc.2025.108963

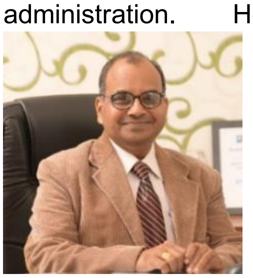
Author Biography

Dr. B. Krishnakumar is working as an Assistant Professor in the School of Computing at SASTRA Deemed University, Tamil Nadu, India. He holds a B.Tech degree in Information Technology and an M.E. degree in Computer Science and Engineering from Anna University. He received his Ph.D. from Anna University, specializing in Deep Learning. He has more than 18 years of teaching experience and has handled a wide range of undergraduate and postgraduate courses in Computer Science. He has published over 12 research articles in international journals, including reputed SCI-indexed journals, and more than 17 papers in international and national conferences. He has also authored two book chapters with internationally reputed publishers such as IGI Global and Wiley.

Manuscript received 15 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 1 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1550

P. Nisha (Nisha Paulraj) is currently working as an Assistant Professor in the Department of Computer Science and Engineering at Dr. N.G.P. Institute of Technology, Coimbatore. She completed her B.E. in Computer Science and Engineering from PSR Engineering College, affiliated with Anna University, Thirunelveli, in 2011. She subsequently earned her M.E. degree from the prestigious PSG College of Technology, Coimbatore. She is presently pursuing her Ph.D. in the medical domain, focusing on the application of deep learning techniques for healthcare and biomedical data analysis. Her research interests primarily include Machine Learning, Deep Learning, medical image analysis, and intelligent decision support systems for clinical applications.

Dr. Sri Laxmi K is an Associate Professor in the Department of Computer Science and Engineering at CVR Engineering College with over 24 years of teaching and research experience in higher education. She earned her Ph.D. in Computer Science and Engineering. Her research interests span deep learning, medical image analysis, machine learning, data analytics, big data analytics, and intelligent systems. She has published more than 15 research works, including articles in SCIE and SCOPUS-indexed journals, international conferences, book chapters, and patents. She actively serves as a reviewer for reputed international journals and IEEE conferences and has contributed as a resource person and coordinator for faculty development programs, workshops, conferences, boot camps, and hackathons. She has played a significant role in curriculum development, research mentoring, and academic administration. Her sustained contributions to teaching, research, and interdisciplinary collaboration reflect her commitment to advancing engineering education and impactful research.



Dr. K. Venu is currently serving as an Assistant Professor (Senior Grade) in the Department of Computer Science and Engineering at Kongu Engineering College, Tamil Nadu, India. She completed her Ph.D. in Deep Learning at Anna University, where her research focused on advanced learning models and intelligent computing techniques. With over 9 years of teaching experience, she has actively contributed to both academic instruction and research mentoring. She has published 7 research articles in reputed international journals and presented 11 papers at international and national conferences. In addition, she has authored 3 book chapters with well-known publishers. Her research interests include Machine Learning, Deep Learning, and Nature-Inspired Computing, with applications in intelligent and data-driven systems.

Mrs. R. Evance Leethial is currently serving as an Assistant Professor in the Department of Computer Science and Engineering at Nehru Institute of Technology, Coimbatore. She holds a Bachelor's degree in Information Technology and a Master's degree in Computer Science and Engineering, reflecting her strong academic foundation in the field of computing. With a deep commitment to teaching and learning, she has been actively involved in academia and curriculum development, contributing to her students' academic growth through innovative teaching methods and ongoing mentorship. She has a keen interest in Artificial Intelligence (AI) and Machine Learning (ML), and has made significant contributions through research and scholarly activities. She has presented six research papers in reputed national and international conferences, and has successfully published her work in foreign journals, showcasing her engagement with the global research community. Her areas of expertise include supervised learning, data-driven modeling, deep learning, and AI-based intelligent systems. In addition to teaching and research, she also participates in academic events, faculty development programs, and technical workshops, thereby enriching her knowledge and staying updated with emerging trends in the field of computer science.

Dr. Kunchanapalli Rama Krishna is a Professor in the CSIT Department at KL University with over 33 years of academic and research experience in Computer Science & Engineering. He holds a Ph.D. in

Manuscript received 15 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 1 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1550

CSE and has served at reputed institutions, including Galgotias University, C-DAC, and IIMT College of Engineering. As an AICTE Margdarshak and NAAC e-Assessor, he has guided institutions toward NBA accreditation and academic excellence. He has published over 30 research papers in reputed national and international journals and conferences. Dr. Rama Krishna has also led funded projects and national conferences on cryptography and modernizing education. He actively participates in FDPs and holds certifications from IITs, IIMs, and NPTEL/SWAYAM. His guidance has shaped numerous UG, PG, and Ph.D. scholars, contributing to quality education and innovation.

Manuscript received 15 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 1 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1550