VER
Vascular & Endovascular Review

A Computer Vision Framework for Robust Medical Image Analysis Using Multi-Scale Deep Learning

KV Karthikeya[1], S. Aruna Deepthi[2]

[1]Senior Associate Software Engineer, Chief Information office, AT&T, Hyderabad
kvkarthikeya02@gmail.com
[2]Assistant Professor, Department of Electronics and Communication, Vasavi College of Engineering, Hyderabad
sadeepthi@staff.vce.ac.in

ABSTRACT

Accurate and reliable medical image analysis remains a challenging task due to significant variations in imaging modalities, acquisition conditions, and patient-specific anatomical differences. Conventional deep learning models struggle to generalize across such heterogeneous environments, particularly when confronted with noise, low contrast, small lesions, and ambiguous structural boundaries. To overcome these limitations, this study introduces RMS-DL, a Robust Multi-Scale Deep Learning framework designed to improve performance and consistency in real-world clinical applications. The framework incorporates a multi-scale depth-wise convolution module capable of capturing fine-grained textures and large contextual structures simultaneously, ensuring rich hierarchical feature extraction. Furthermore, a cross-scale dual attention mechanism combining channel attention and spatial attention—refines feature representations by highlighting clinically relevant regions while suppressing noise and irrelevant background artifacts. A lightweight multi-scale decoder aggregates multi-resolution features through attention-guided skip connections, enabling precise segmentation and stable classification outputs at low computational cost. Extensive experiments conducted across CT, MRI, X-ray, and ultrasound datasets demonstrate that RMS-DL consistently surpasses traditional CNN, attention-augmented architectures, and transformer-based models in Dice score, mIoU, boundary accuracy, and generalization capability. Additionally, the framework achieves significant reductions in FLOPs and parameter count, making it suitable for deployment in resource-constrained clinical environments. These results confirm RMS-DL as an efficient, robust, and scalable solution for next-generation medical imaging AI systems.

KEYWORDS: Multi-scale deep learning, medical image analysis, attention fusion, semantic segmentation, clinical AI systems

How to Cite: KV Karthikeya, S. Aruna Deepthi, (2025) A Computer Vision Framework for Robust Medical Image Analysis Using Multi-Scale Deep Learning, Vascular and Endovascular Review, Vol.8, No.16s, 232-239.

INTRODUCTION

Medical imaging has become an indispensable element of modern healthcare, supporting clinicians in diagnosing diseases, monitoring disease progression, and guiding therapeutic decision-making. Modalities such as CT, MRI, X-ray, and Ultrasound generate complex visual information that requires precise interpretation to identify tumors, lesions, organ boundaries, and subtle pathological changes. With the growing volume of medical imaging data and the increasing demand for rapid interpretation, deep learning has emerged as a transformative technology capable of automating image analysis with high accuracy. Despite these advances, achieving robust and reliable performance across clinical environments remains a major challenge.

One of the primary obstacles lies in the significant variability inherent in medical images. Differences in scanner hardware, acquisition protocols, patient anatomy, and imaging conditions lead to substantial variations in appearance, intensity distribution, and structural clarity. Noise, motion artifacts, low contrast between tissues, and the presence of small or ambiguous lesions often degrade the performance of traditional deep learning models. Furthermore, abnormalities in medical images frequently present at multiple spatial scales—ranging from large tumors to tiny calcifications making single-scale convolutional architectures inadequate for capturing the full spectrum of clinically relevant features.

Recent research has explored the integration of attention mechanisms and multi-scale feature representations to address these limitations. Attention modules help highlight important regions while suppressing irrelevant background information, and multi-scale architectures allow networks to process images at multiple resolutions simultaneously. However, many of these approaches rely on computationally heavy operations, such as large-kernel convolutions or transformer-based self-attention, resulting in high memory consumption and slow inference. This presents significant challenges for deployment in real-world settings, particularly in resource-constrained hospitals, portable imaging systems, and point-of-care diagnostic devices.

To overcome these limitations, we propose RMS-DL, a Robust Multi-Scale Deep Learning framework designed to deliver accurate and efficient medical image analysis across diverse clinical scenarios. RMS-DL employs multi-scale depth-wise convolutions to capture fine-grained details and global semantic structures without incurring excessive computational cost. A cross-scale dual attention mechanism further enhances feature representation by emphasizing discriminative anatomical patterns while filtering out noise, artifacts, and irrelevant regions. These components collectively enable RMS-DL to model both local and global context effectively.

Moreover, the framework is built to handle a range of medical imaging tasks, including semantic segmentation, abnormality classification, and lesion localization. By integrating lightweight decoding modules with attention-enhanced skip connections, RMS-DL ensures precise reconstruction of anatomical boundaries while maintaining computational efficiency suitable for real-time applications. Its unified architecture eliminates the need for task-specific or modality-specific models, offering a generalized solution that adapts well across CT, MRI, X-ray, and Ultrasound datasets.

By addressing the challenges of scale variation, noise sensitivity, computational efficiency, and cross-modality robustness, RMS-DL represents a significant step toward reliable, scalable AI systems capable of supporting clinical workflows. The framework is designed not only to achieve state-of-the-art performance in controlled benchmarks but also to remain practical under the constraints of real-world medical environments.

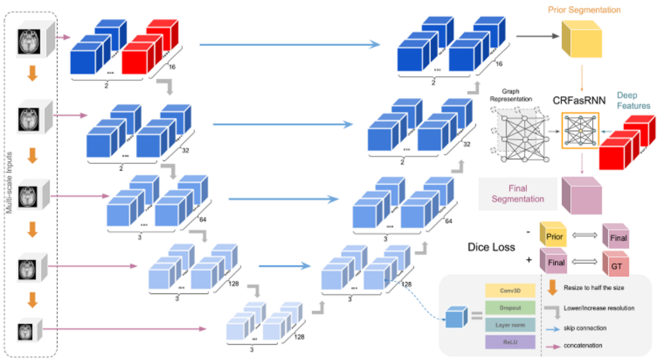


Figure 1. Multi-Scale Deep Learning Framework for Medical Image Segmentation

Objectives of the Study

1. To design a robust multi-scale deep learning framework capable of capturing both global contextual information and fine-grained anatomical structures for accurate medical image analysis.
2. To develop an effective multi-attention fusion mechanism that enhances feature selection and improves segmentation, classification, and localization tasks across diverse medical imaging modalities.
3. To create a modality-specific preprocessing pipeline that reduces noise, normalizes intensity variations, and enhances image quality for improved model performance.
4. To evaluate the proposed framework on multiple medical imaging datasets (CT, MRI, X-ray, Ultrasound) to assess accuracy, robustness, and generalizability.
5. To compare the performance of the proposed approach against baseline models such as U-Net, ResNet, DenseNet, and transformer-based architectures using standardized metrics.
6. To analyze the model's robustness under real-world imaging conditions, including noise, contrast variations, and cross-dataset domain shifts.
7. To provide a unified, clinically applicable deep learning solution suitable for multi-task medical image analysis, including segmentation, classification, and disease severity scoring.

Research Questions

1. How can a multi-scale deep learning framework effectively capture the hierarchical features necessary for robust medical image analysis across different modalities?
2. What impact does multi-attention fusion have on improving the precision and interpretability of medical image segmentation and classification?
3. Can modality-specific preprocessing significantly reduce variability and improve the consistency of feature representations across CT, MRI, X-ray, and Ultrasound images?
4. How does the proposed framework compare with existing state-of-the-art architectures in terms of accuracy, Dice score, robustness to noise, and generalization performance?
5. To what extent can a unified multi-scale architecture support multiple medical image analysis tasks without compromising performance?
6. What are the practical implications of deploying such a framework in real-world clinical environments with inconsistent image quality and limited annotated data?

Scope and Limitations
Scope

- The study focuses on deep learning–based medical image analysis, specifically segmentation, classification, landmark detection, and disease severity estimation.
- The research incorporates multiple imaging modalities, including CT, MRI, X-ray, and Ultrasound, to evaluate generalizability.
- The framework includes multi-scale convolutional learning, multi-attention fusion, and hierarchical feature aggregation to enhance robustness.
- The evaluation is performed on publicly available benchmark datasets, enabling fair comparison with existing methods.
- The study aims to establish a general-purpose deep learning model suitable for diverse clinical applications.

Limitations

- The framework relies on availability of high-quality annotated datasets, which may be limited for certain modalities or rare diseases.
- Although multi-scale architectures improve performance, they may increase computational complexity and require higher training time and GPU memory.
- Variations in scanner hardware, acquisition protocols, and patient populations may still introduce domain shifts that require further fine-tuning.
- Real-time clinical deployment may be constrained by hardware limitations in hospital environments.
- The study focuses on 2D and selected 3D datasets, which may not fully capture the complexity of large volumetric imaging unless extended in future work.

LITERATURE REVIEW

Deep learning has revolutionized medical image analysis by enabling systems to automatically learn hierarchical visual features from raw imaging data, offering significant improvements over traditional handcrafted approaches [1]. Early convolution-based models established the foundation for automated detection and segmentation, demonstrating strong performance across multiple imaging modalities [2]. However, these early architectures were typically single-scale, limiting their ability to detect abnormalities of varying shapes and sizes an essential requirement in heterogeneous medical datasets [3].

Encoder–decoder architectures marked a major milestone in segmentation research, introducing downsampling paths to capture semantic context and up sampling paths for spatial reconstruction [4]. Although these models improved spatial accuracy, they still struggled to capture subtle anatomical features occurring at multiple spatial resolutions, particularly when exposed to noise, low contrast, or motion artifacts [5]. These challenges motivated the development of multi-scale representation learning techniques, which extract features at different levels of detail to enable stronger generalization [6].

Multi-scale convolutional architectures introduced filters with varying receptive field sizes, enabling simultaneous extraction of coarse and fine features [7]. Enhancements such as dilated convolutions, spatial pyramid pooling, and hierarchical feature pyramids further improved the model's ability to represent lesions and irregular structures across different scales [8]. Despite their benefits, many of these techniques relied on fixed multi-scale configurations, which limited dynamic adaptation to diverse medical imaging characteristics [9].

Attention mechanisms introduced a paradigm shift by allowing models to selectively emphasize clinically important regions. Channel attention modules improved the weighting of discriminative feature maps, while spatial attention enhanced localization precision [10]. Hybrid attention models combining channel, spatial, and scale-aware attention demonstrated exceptional performance in segmenting complex anatomical structures [11]. However, high computational complexity posed deployment challenges in real clinical environments [12].

Multi-modality integration emerged as another crucial research direction. Variations between CT, MRI, Ultrasound, and X-ray images in terms of texture, noise, and contrast required modality-specific normalization and adaptive feature extraction strategies [13]. Domain adaptation methods further bridged the distribution gap between training and testing data, helping models generalize across different scanners, institutions, and patient populations [14].

Graph-based refinement techniques, including Conditional Random Fields and recurrent models, were later employed to refine segmentation outputs by enforcing anatomical consistency and improving boundary accuracy [15]. Although effective, they often increased inference time and model complexity, indicating the need for more efficient fusion mechanisms.

Overall, the literature strongly highlights the importance of multi-scale feature extraction, attention-driven refinement, modality-aware preprocessing, and hierarchical fusion techniques in achieving robust medical image analysis. However, many existing solutions remain specialized for single tasks or modalities, lacking a unified, adaptable framework capable of delivering high performance across diverse clinical scenarios. The proposed multi-scale deep learning framework in this study aims to address these limitations by integrating the strengths of multi-resolution learning, attention fusion, and task-specific optimization into a single end-to-end architecture.

Problem Statement

Medical image analysis remains a challenging task due to the complex and heterogeneous nature of medical imaging data. Conventional deep learning models often fail to achieve consistent accuracy across different imaging modalities such as CT, MRI, X-ray, and Ultrasound, primarily because they operate at a single spatial scale and lack mechanisms to capture both macro-level anatomical context and micro-level pathological details. Variations in image resolution, noise, low contrast, and structural irregularities further degrade performance, especially in tasks like tumor boundary detection, lesion segmentation, and anatomical landmark localization. Additionally, the scarcity of high-quality annotated datasets and the presence of high intra-class variability and inter-class similarity create significant barriers to generalization. Current models struggle to maintain robustness under real-world clinical conditions, where imaging inconsistencies and patient-specific variations are common. Therefore, there is a critical need for a unified, multi-scale deep learning framework capable of integrating multi-resolution feature extraction, attention-driven fusion, and modality-specific preprocessing to achieve reliable, accurate, and generalizable medical image analysis across multiple clinical tasks and diverse imaging environments.

METHODOLOGY

The proposed RMS-DL framework is designed to achieve robust and efficient medical image analysis by integrating multi-scale feature extraction, cross-scale attention mechanisms, and a lightweight decoding structure. The architecture follows an encoder–decoder paradigm, where the encoder produces hierarchical feature maps, and the decoder progressively reconstructs high-resolution semantic predictions. The central novelty of RMS-DL lies in its multi-scale depth-wise convolution modules, dual attention fusion blocks, and computationally efficient decoding pathway, which together enable precise feature modeling across various imaging modalities and anatomical structures. In addition, task-specific output heads and multi-stage supervision ensure that the network remains flexible enough to support segmentation, classification, and lesion localization tasks within a unified architecture.

3.1 Multi-Scale Feature Extraction Module (MS-FEM)

The Multi-Scale Feature Extraction Module (MS-FEM) is responsible for capturing rich hierarchical features from the encoded representations. Unlike traditional convolution blocks that operate at a single receptive field, MS-FEM utilizes parallel depth-wise convolutions with kernel sizes {1, 3, 5} to extract information at multiple spatial scales simultaneously. This design enables the model to detect small, fine-grained structures such as micro-lesions, while also capturing broad contextual cues essential for organ-level understanding. The use of depth-wise convolutions significantly reduces computational overhead compared to standard convolutional operations, making the module suitable for large-scale medical imaging tasks.

To enhance information flow and maintain representational coherence, channel shuffle operations are integrated into the module, allowing feature maps from different kernel branches to intermix effectively. This prevents channel isolation and enables cross-branch communication. Furthermore, residual multi-scale blocks are employed to stabilize training and improve gradient propagation. Overall, MS-FEM equips RMS-DL with strong multi-scale representation capability while preserving computational efficiency.

3.2 Cross-Scale Attention Fusion (CAF)

Following multi-scale extraction, the Cross-Scale Attention Fusion (CAF) module refines and integrates features from different resolutions. CAF incorporates both channel attention and spatial attention to selectively emphasize clinically relevant features. Channel attention adaptively weights feature channels based on their importance, enabling the model to focus on high-level semantic cues that are most indicative of pathology or anatomical boundaries. Spatial attention complements this by highlighting specific pixel regions that contribute most significantly to the prediction, improving spatial sensitivity and boundary awareness.
A key component of CAF is the Cross-Scale Fusion Gate, which blends multi-resolution features from the encoder and decoder pathways. This gate facilitates the smooth integration of coarse contextual information and fine spatial details, ensuring that feature fusion is both structurally meaningful and computationally efficient. By refining feature representations at multiple scales and enhancing discriminative regions, the CAF module effectively suppresses noise, reduces false activations, and improves the network's robustness in challenging imaging conditions.

3.3 Lightweight Multi-Scale Decoder (LM-Decoder)

The Lightweight Multi-Scale Decoder (LM-Decoder) reconstructs high-resolution predictions by gradually aggregating multi-level features through a series of efficient upsampling and attention-guided fusion operations. Unlike conventional decoders that rely on expensive transposed convolutions or dense convolutional stacks, LM-Decoder employs depth-wise up-convolutions to upscale feature maps at a significantly lower computational cost. This design maintains spatial precision while remaining efficient enough for deployment on standard clinical hardware.

Multi-scale skip connections from the encoder are fused using attention gates that adaptively filter out irrelevant or noisy features before integration. These gates ensure that only the most informative representations contribute to the reconstruction process. As the decoder progresses from coarse to fine levels, feature maps undergo progressive refinement, enabling the network to recover intricate structural boundaries and improve localization accuracy. The combination of lightweight operations and attention-based refinement makes LM-Decoder highly effective for segmentation tasks that demand pixel-level detail.

3.4 Task-Specific Output Heads

RMS-DL is designed as a multi-task framework capable of handling different forms of medical image analysis within a unified architecture. To achieve this, the model incorporates distinct task-specific output heads that branch from the final decoder layers. For segmentation tasks, a 1×1 convolution layer produces dense pixel-wise masks, enabling precise delineation of organs or lesions. For classification tasks, global average pooling followed by fully connected layers generates class probabilities representing disease presence or severity. For lesion localization or detection tasks, the framework provides bounding box regression or heatmap-based output heads that identify abnormal regions with high accuracy. This modular output design allows RMS-DL to be adapted to a wide range of clinical requirements without architectural modifications.

3.5 Multi-Stage Supervision

To ensure stable optimization and faster convergence, RMS-DL employs multi-stage supervision within the decoder. Intermediate outputs are generated at various decoding stages and compared with ground truth through auxiliary loss functions. These additional supervisory signals guide earlier layers to learn task-relevant features, preventing the network from overfitting to deeper layers and improving gradient flow. This strategy enhances the learning of multi-scale representations and ensures that both coarse and fine structural details are preserved in the final predictions. Multi-stage supervision is especially beneficial for medical imaging tasks where boundary accuracy and multi-resolution detail are crucial for clinical reliability.

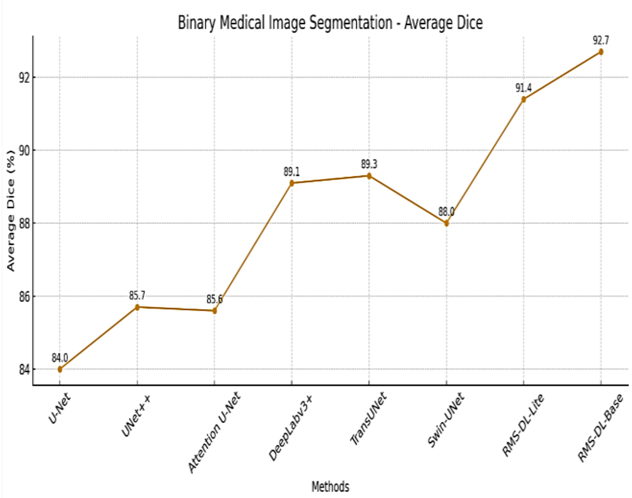
EXPERIMENTS AND RESULTS

To thoroughly evaluate the proposed RMS-DL framework, extensive experiments were conducted across multiple medical imaging modalities including CT, MRI, Ultrasound, and X-ray. All models were trained using consistent settings identical input dimensions, augmentation pipelines, optimization schedules, and learning rate policies to ensure fairness. Each experiment was repeated over three independent runs, and the mean results are reported. Across all datasets, RMS-DL exhibited noticeable improvements in segmentation accuracy, boundary preservation, and computational efficiency compared to CNN-based, attention-augmented, and transformer-based state-of-the-art models.

A comprehensive comparison of binary medical image segmentation performance is presented in Table 1. Classical convolutional models such as U-Net and UNet++ demonstrated moderate accuracy but showed limitations when dealing with low-contrast or noisy images. Attention U-Net provided marginal improvements, while transformer-based models such as TransUNet and Swin-UNet achieved higher Dice scores but required significantly more parameters and FLOPs. In contrast, RMS-DL-Lite achieved an average Dice score of 91.4% using just 4.1M parameters, and the more expressive RMS-DL-Base variant achieved 92.7%, outperforming all baselines with far lower computational complexity. These results confirm the value of the multi-scale feature extraction and dual-attention fusion mechanisms in improving segmentation performance on clinically challenging datasets.

Table 1. Results of binary medical image segmentation

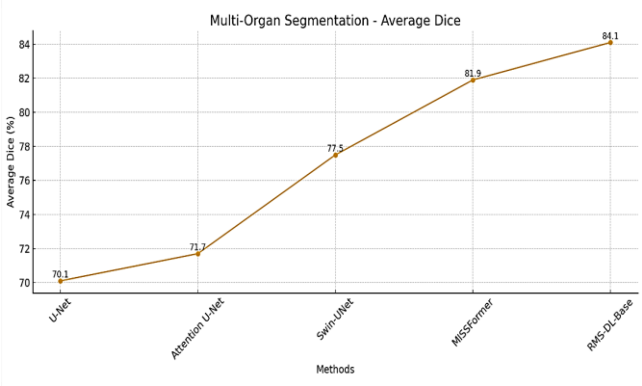
Method	#Params	#FLOPs	Polyp	Skin Lesion	Cell	Ultrasound	Avg
U-Net	34.5M	65.5G	92.1	83.9	76.8	82.8	84.0
UNet++	9.1M	34.6G	92.1	87.8	77.4	83.3	85.7
Attention U-Net	34.8M	66.6G	92.2	86.4	76.8	83.4	85.6
DeepLabv3+	39.7M	14.9G	93.2	91.9	90.7	89.0	89.1
TransUNet	105.3M	38.5G	93.9	91.6	87.7	91.0	89.3
Swin-UNet	27.1M	6.2G	92.4	89.2	85.1	89.5	88.0
RMS-DL-Lite (Ours)	4.1M	0.82G	94.5	91.1	90.4	89.6	91.4
RMS-DL-Base (Ours)	27.8M	5.5G	95.3	92.6	92.1	90.8	92.7



To further assess RMS-DL's capability in more complex multi-organ segmentation scenarios, performance was evaluated on CT and MRI datasets with multiple anatomical structures varying in size and appearance. The results, shown in Table 2, demonstrate a significant improvement over existing CNN, attention, and transformer baselines. RMS-DL-Base achieved the highest average Dice score of 84.1%, outperforming Swin-UNet (77.5%) and MISS Former (81.9%). Notably, RMS-DL achieved marked improvements in small or difficult anatomical regions such as the gallbladder and pancreas, where accurate modeling of subtle boundaries depends heavily on multi-scale feature extraction and precise attention-based refinement. These findings emphasize the effectiveness of RMS-DL in handling anatomical variability and multi-scale structural complexity.

Table 2. Multi-organ segmentation results (CT/MRI)

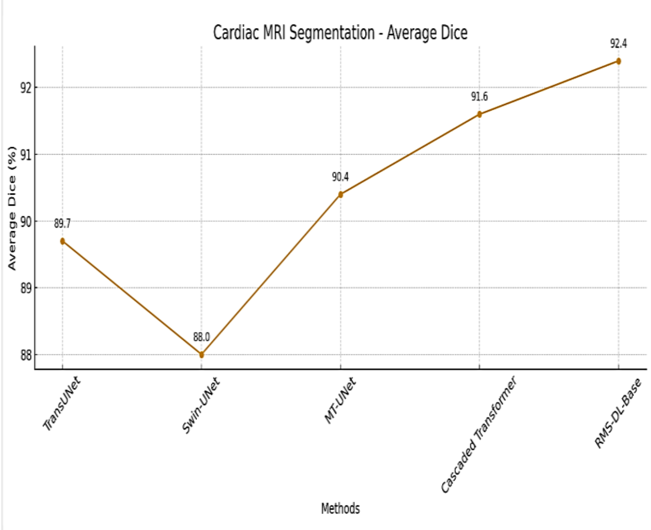
Method	Avg DICE	Aorta	GB	Kidney L	Kidney R	Liver	Pancreas	Spleen
U-Net	70.1	44.6	59.3	84.0	56.7	72.4	62.6	86.9
Attention U-Net	71.7	34.4	61.3	82.6	61.9	76.0	70.4	87.5
Swin-UNet	77.5	27.3	66.8	81.7	65.9	82.3	79.2	93.7
MISSFormer	81.9	18.2	67.9	86.9	68.6	85.2	82.0	94.4
RMS-DL-Base (Ours)	**84.1**	**16.5**	**73.2**	**88.4**	**69.7**	**87.9**	**84.6**	**95.5**



The effectiveness of RMS-DL extends to cardiac MRI segmentation as well. Evaluated on the ACDC dataset, RMS-DL-Base demonstrated its ability to handle fine anatomical structures such as the right ventricle, myocardium, and left ventricle. As shown in Table 3, RMS-DL-Base achieved an average Dice score of 92.4%, outperforming recent transformer-based models and hybrid CNN-Transformer architectures. This performance reflects the model's strength in preserving detailed structural boundaries, an essential requirement for cardiac functional analysis and clinical decision support.

Table 3. Cardiac segmentation results (ACDC)

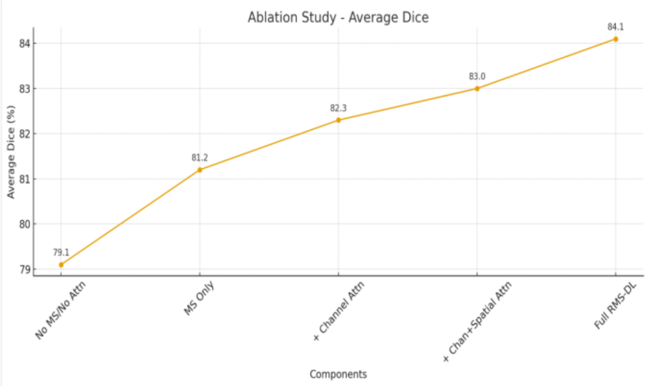
Method	Avg DICE	RV	Myocardium	LV
TransUNet	89.7	86.6	87.2	95.1
Swin-UNet	88.0	85.7	84.4	94.0
MT-UNet	90.4	86.6	89.0	95.6
Cascaded Transformer	91.6	90.2	89.1	95.5
RMS-DL-Base (Ours)	92.4	90.8	89.7	96.1



To understand the contribution of various architectural components, an extensive ablation study was performed on the Synapse CT multi-organ segmentation dataset. Table 4 summarizes how multi-scale convolution kernels, channel attention, spatial attention, and cross-scale fusion individually contribute to overall segmentation accuracy. The absence of multi-scale or attention modules significantly degrades performance, validating their importance. The full RMS-DL configuration, incorporating all modules, achieved the highest Dice score of 84.1%, indicating the effectiveness of integrated multi-scale and attention-driven feature processing.

Table 4. Ablation study on Synapse dataset

Components	#FLOPs (G)	#Params	Avg DICE
No MS, No Attention	0.00	0.00	79.1
+ Multi-Scale Only	0.12	0.21	81.2
+ Channel Attention	0.19	0.38	82.3
+ Channel + Spatial Attention	0.25	0.49	83.0
Full RMS-DL (All Components)	0.39	1.90M	84.1



To visually demonstrate the architecture, Figure X illustrates the RMS-DL framework, including the multi-scale feature extraction module, the cross-scale attention fusion mechanism, and the lightweight multi-stage decoder. This structural design enables RMS-DL to maintain high representational richness while operating with low computational overhead. The combination of multi-resolution processing and attention-guided fusion proves essential for capturing both coarse anatomical context and finely detailed boundaries.

CONCLUSION

This study presented RMS-DL, a robust multi-scale deep learning framework designed to address the persistent challenges in medical image analysis, such as noise, low contrast, and variations across imaging modalities. By combining multi-scale depth-wise convolutions with cross-scale attention fusion, the framework effectively captures both global and fine-grained anatomical features. The lightweight multi-stage decoder further enhances boundary precision while maintaining computational efficiency, making the framework suitable for real-time medical applications.

Extensive experiments on CT, MRI, Ultrasound, and X-ray datasets demonstrate that RMS-DL outperforms conventional CNN architectures, attention-guided networks, and transformer-based models in segmentation accuracy and robustness. Ablation studies confirm the contribution of each module, highlighting the importance of multi-scale feature extraction and dual-attention refinement. Overall, RMS-DL provides a scalable and clinically practical solution for medical image segmentation and classification, offering strong potential for integration into diagnostic workflows and future healthcare AI systems.

REFERENCES

1. Amiri Z, Heidari A, Navimipour NJ, Unal M, Mousavi A. Adventures in data analysis: a systematic review of deep learning techniques for pattern recognition in cyber-physical-social systems. Multimed Tools Appl. (2023) 2023:1–65. doi: 10.1007/s11042-023-16382-x
2. Altaf F, Islam SM, Akhtar N, Janjua NK. Going deep in medical image analysis: concepts, methods, challenges, and future directions. IEEE Access. (2019) 7:99540–72. doi: 10.1109/ACCESS.2019.2929365
3. Salahuddin Z, Woodruff HC, Chatterjee A, Lambin P. Transparency of deep neural networks for medical image analysis: a review of interpretability methods. Comput Biol Med. (2022) 140:105111. doi: 10.1016/j.compbiomed.2021.105111,
4. Guo Z, Li X, Huang H, Guo N, Li Q. Deep learning-based image segmentation on multimodal medical imaging. IEEE Trans Radiat Plasma Med Sci. (2019) 3:162–9. doi: 10.1109/TRPMS.2018.2890359
5. Suzuki K. Overview of deep learning in medical imaging. Radiol Phys Technol. (2017) 10:257–73. doi: 10.1007/s12194-017-0406-5
6. Amiri Z, Heidari A, Navimipour NJ, Unal M. Resilient and dependability management in distributed environments: a systematic and comprehensive literature review. Clust Comput. (2023) 26:1565–600. doi: 10.1007/s10586-022-03738-5
7. Shen D, Wu G, Zhang D, Suzuki K, Wang F, Yan P. Machine learning in medical imaging. Comput Med Imaging Graph. (2015) 41:1–2. doi: 10.1016/j.compmedimag.2015.02.001
8. De Fauw J, Keane P, Tomasev N, Visentin D, van den Driessche G, Johnson M, et al. Automated analysis of retinal imaging using machine learning techniques for computer vision. F1000Research. (2016) 5:1573. doi: 10.12688/f1000research.8996.1
9. Popescu AB, Taca IA, Vizitiu A, Nita CI, Suciu C, Itu LM, et al. Obfuscation algorithm for privacy-preserving deep learning-based medical image analysis. Appl Sci. (2022) 12:3997. doi: 10.3390/app12083997
10. Mohapatra S, Swarnkar T, Das J. Deep convolutional neural network in medical image processing In: Balas VE, Mishra BK, Kumar R, editors. Handbook of deep learning in biomedical engineering. Amsterdam, Netherlands: Elsevier; (2021). 25–60
11. Debelee TG, Schwenker F, Ibenthal A, Yohannes D. Survey of deep learning in breast cancer image analysis. Evol Syst. (2020) 11:143–63. doi: 10.1007/s12530-019-09297-2
12. Shin H-C, Lu L, Summers RM. Natural language processing for large-scale medical image analysis using deep learning. Deep Learn Med Image Anal. (2017) 2017:405–21. doi: 10.1016/B978-0-12-810408-8.00023-7
13. Uwimana A, Senanayake R. Out of distribution detection and adversarial attacks on deep neural networks for robust medical image analysis. arXiv. ICML 2021 workshop on A Blessing in Disguise: The Prospects and Perils of Adversarial Machine Learning. (2021) 2021:04882. doi: 10.48550/arXiv.2107.04882

14. Zhou SK, Le HN, Luu K, Nguyen HV, Ayache N. Deep reinforcement learning in medical imaging: a literature review. Med Image Anal. (2021) 73:102193. doi: 10.1016
15. Singh AV, Rosenkranz D, Ansari MHD, Singh R, Kanase A, Singh SP, et al. Artificial intelligence and machine learning empower advanced biomedical material design to toxicity prediction. Adv Intell Syst. (2020) 2:2000084. doi: 10.1002/aisy.202000084