| Volume 02 | Issue 02 | June 2026 | ISSN: 3139-1737

International Journal of Advances and Applications in Convergent Studies & Transdisciplinary (IJ-AACST)

Bridging Disciplines, Bridging Knowledge

Aligned with the Vision of NEP 2020 and UGC Guidelines

Website: https://aacstij.org || **Email:** journaleditor@aacstij.org

A Review on Deep Learning Techniques for Medical Image Segmentation and Classification

Ms. Kavita Gawhade[1], Dr. Girish Katkar[2]

[1]Taywade College, Nagpur, Maharashtra, India,
kavita.sfscollege13@gmail.com
ORCID: https://orcid.org/0009-0002-4848-3833

[2]Taywade College, Nagpur, Maharashtra, India,
gkatkar13@gmail.com
ORCID: https://orcid.org/0009-0007-1561-8128

ABSTRACT

Modern healthcare is greatly supported by medical image analysis. For accurate diagnosis and effective treatment planning, Deep Learning techniques have emerged as a powerful approach. Convolutional Neural Network (CNNs) such as U-Net and its variants have gained significant attention for medical image Segmentation and Classification. These models are widely used due to their ability to extract local and global features from medical images.

This paper presents a comprehensive review of various Deep Learning techniques for Medical medical image Segmentation and Classification. Various U-Net variants like attention U-Net, ResUNet, and UNet++ as well as transformer-based models like TransUNet, and 3D models such as V-Net a 3D-UNet for volumetric data analysis are discussed in this paper. The paper also provides a brief analysis of strength, limitations and advancements of various models.

This paper focuses on developing an integrated framework that combines Segmentation and Classification with Attention mechanism and Explainable AI(XAI) and also addresses limitations such as high computational complexity, limited availability of annotated data and increased resource requirements.

KEYWORDS: Medical image Segmentation, Deep Learning, U-Net, Convolutional Neural Networks, Attention Mechanism, Explainable AI (XAI).

1. INTRODUCTION

Medical image analysis plays an important role in modern healthcare and enables early detection of abnormalities in human organs, which leads to accurate diagnosis and effective treatment planning. A large volume of data is generated by Medical imaging techniques such as Computed Tomography (CT scans), Magnetic Resonance Imaging (MRI) and X-Rays imaging. Previously, Medical image analysis performed by experts was time consuming, inefficient for large scale data and prone to inter-observer variability. With the rapid development of Convolutional Neural Networks (CNNs) Deep Learning models have demonstrated superior performance over the time in tasks such as Medical image Segmentation and Classification by learning hierarchical features from data.

The U-Net architecture is an influential and highly efficient technique for detecting abnormalities in medical images. Several variants and advancements have been proposed to enhance performance in image Segmentation and Classification. This includes Dense Skip connections and Residual Learning for improved training efficiency and to reduce semantic gap. Additionally, Attention mechanisms to focus on relevant regions and Transformer based models to capture global contextual information. Moreover, 3D architectures have been developed to handle volumetric medical data.

2. CONTRIBUTION OF THE PAPER

1. This paper presents a comprehensive review of Deep Learning methods for image Segmentation and Classification techniques. The Deep Learning techniques covered by this paper are U-Net, UNet++, ResUNet, Attention U-Net, TransUNet , V-Net and 3D-UNet. It provides the comparative analysis of their architectures, strengths, limitations and suitability for various Medical imaging databases.

2. This paper identifies the Research gap, limitation of integrated approach Segmentation and Classification algorithms, limited annotated medical databases, high computational cost, and lack of explainability. Based on these interpretations this paper proposes an integrated framework for Medical image Segmentation and Classification, incorporating Attention Gates and Explainable AI(XAI).

3. LITERATURE REVIEW / RELATED WORK / BACKGROUND STUDY

1. U-Net and basic Segmentation models

1.1 U-Net Architecture **[1]** Ronneberger et al. "U-Net: Convolutional Networks for Biomedical Image Segmentation," MICCAI, 2015.

The U-Net architecture is based on encoder decoder framework with Skip connections that enables precise localization of lesions or tumors while preserving important contextual information. Encoder path captures high level semantic features and decoder path reconstructs segmented output with improved spatial accuracy. The Skip connections help transferring fine grained feature information from encoder to decoder. U-Net is very efficient in learning effectively from limited annotated datasets, which makes it highly suitable for biomedical applications where large scale labeled data is often unavailable. U-Net has become a widely used foundation architecture for Medical image Segmentation and has inspired many variants over the time.

1.2 SegNet [2] Badrinarayanan V.et al. "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2017.

A deep Convolutional encoder decoder architecture was introduced for image Segmentation tasks which utilize pooling indices during the up-sampling process which helps in reducing memory consumption and improving computational efficiency. The architecture performs Semantic Segmentation with very few trainable parameters and in faster inference time. SegNet generally demonstrate lower Segmentation accuracy as compared to UNet in complex medical imaging application due to the limited transfer of spatial information during decoding.

2. Improved UNet Variants

2.1 UNet++ [3] Zhou Z. et al. "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," Deep Learning in Medical Image Analysis (DLMIA) Workshop, 2018.

UNet++ is a nested U-Net architecture for Medical image Segmentation which introduces nested and Skip connections to reduce the semantic gap between encoder and decoder feature maps. These redesigned Skip pathways facilitate better feature fusion and improved information propagation throughout the network. As a result, the model achieves enhanced feature representation, improved Segmentation accuracy and better localization performance compared to the original U-Net architecture.

2.2 Residual UNet [4] Zhang Z. et al."Road Extraction by Deep Residual U-Net," IEEE Geoscience and Remote Sensing Letters, 2018

Residual U-Net (ResUNet) utilizes Residual connections within Convolutional layers to overcome the Vanishing gradient problems and improve model convergence during training. The Residual connections enable better feature propagation and allow the network to learn complex representations more effectively. ResUNet improve Segmentation accuracy and training stability as compared to U-Net architecture.

2.3 nnU-Net [5] Isensee F. et al. "nnU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation," Nature Methods, 2021.

nnU-UNet is a Self-adapting framework for U-Net based Medical image Segmentation facilitates the design and configuration of Segmentation pipelines by adapting network parameters according to different datasets. nnU-Net automatically selects pre-processing steps training strategies and architectural configuration without recovering extensive manual tuning. Self-adapting capability of this framework enables state-of-the-art performance across multiple Medical image Segmentation tasks and also reduces the dependency on expert intervention.

3. Attention - Based Models [6] Oktay O.et al. "Attention U-Net: Learning Where to Look for the Pancreas," Medical Imaging with Deep Learning (MIDL), 2018.

Attention U-Net uses Attention mechanisms in Deep Learning architectures to improve the ability of Segmentation models and to focus on relevant regions in medical images. The integrated Attention Gates into the traditional U-Net architecture enables the network to suppress irrelevant background features and emphasize important anatomical structures. Attention gates also help the model to identify and focus on target regions of varying sizes and shapes without additional localization models. In complex Medical imaging tasks involving small organs or low-contrast regions, the Attention mechanism highlights useful features during the decoding process and has improved Segmentation accuracy and sensitivity. It also facilitates minimal computational overhead while significantly enhancing the overall performance.

4. Transformer-Based Models [7] Chen J. et al. "TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation," arXiv, 2021.

TransUNet combines the strength of CNNs and Transformers to capture both local and global contextual information from medical images. In this architecture, CNNs extract detailed spatial and low-level features and Transformers utilize self-attention mechanism and model long range dependencies and global relationships within the image. Due to this combination TransUNet achieves improved Segmentation accuracy and enhanced feature representation as compared to old CNN models. The integration of Transformers also increases memory usage, training requirements and computational complexity.

5. 3D Segmentation Models

5.1 V-Net [8] Milletari F. et al. "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," Fourth International Conference on 3D Vision (3DV), 2016.

V-Net fully Convolutional Neural Network for volumetric Medical image Segmentation is an extension of Convolutional Neural Network to 3D medical data. It also introduces the Dice Loss function to address the issue of class imbalance which is commonly present in medical datasets. V-Net architecture directly learns volumetric features from input data and enables accurate Segmentation of complex anatomical structures.

5.2 3D U-Net [9] Çiçek Ö. et al. "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," MICCAI, 2016.

3D U-Net learns Dense volumetric Segmentation from sparse annotation and it is an extension of the traditional U-Net architecture to three-dimensional medical images. 3D U-Net replaces 2D operations with their 3D counterparts and useful in medical applications where fully annotated volumetric data is limited.

Due to the processing of 3D data these models generally required high computational power, longer training time and increased memory usage.

4. RESEARCH METHODOLOGY

This literature review shows that Medical image Segmentation has focused on improving the U-Net architecture through various architectural enhancements including Dense Skip pathways, Attention mechanisms, Residual connections and Transformer based encoders.

These modifications have considerably improved the feature representation and Segmentation accuracy. For example, UNet++ performed an average improvement of approximately 2.8 - 3.3% in Intersection over Union (IoU) as compared to traditional U-Net architectures. Attention U-Net enables selective focus on target anatomical regions and improves model sensitivity. TransUNet combines local spatial feature extraction of CNNs with the global contextual understanding provided by Self attention mechanisms.

Despite these advancements, several limitations continue to persist. TransUNet and 3D Segmentation Models involve high computational complexity longer training times and increased GPU memory consumption. It results in resource constrained clinical environments and challenging deployment. Moreover, many Deep Learning models require large annotated datasets which are limited due to privacy constraints and costly expert annotations. There exists a limited focus on integrating Classification tasks with Segmentation for comprehensive disease diagnosis. The limited generalization capability of models is another challenge because the architecture training on one kind of dataset modality can demonstrate reduced performance when applied to another kind of dataset or real-world clinical scenarios.

Comparative Analysis

This section presents a comparative analysis of the reviewed models in terms of computational complexity, Segmentation performance architectural design and its applicability in Medical imaging.

U-Net architecture is a simple yet highly effective encoder-decoder architecture with Skip connections which enables precise localization and efficient training with limited annotated datasets. It also preserves the spatial information during feature reconstruction. U-Net architecture became the foundation for several subsequent Segmentation architectures due to its strong performance and efficiency. However, this model struggles in dealing heterogeneous datasets and capturing complex contextual dependencies. These limitations were addressed by UNet++ which introduced nested and Dense skip pathways and reduced the semantic gap between encoder and decoder feature maps and also improved feature fusion and information propagation resulting in enhanced Segmentation accuracy.

Attention based approaches further enhance Segmentation performance by integrating Attention gates that focuses on relevant anatomical structures and suppresses relevant background information resulting in improved Segmentation sensitivity and localization accuracy.

TransUNet combines CNN with transformer-based encoders to capture both long range contextual dependencies and spatial details. This hybridization demonstrated superior Segmentation performance in several medical data sets. However, transformer-based architectures require large scale datasets. They are computationally intensive and involves increased GPU memory consumption and training time.

VNet and 3D U-Net learn dense volumetric Segmentation from sparse annotation to extent Segmentation to 3D inputs. It also introduced Dice loss function to effectively handle class imbalance. However, their deployment is constrained by increased memory usage and high computational requirement.

SegNet emphasizes computational efficiency by the use of pooling indices for up-sampling hence reduced memory consumption and inference time. Although it is suitable for real time applications and computationally efficient.

The literature review highlights a clear trade-off between computational efficiency and Segmentation accuracy across different architectures.

Research Gap

Despite significant improvement in Segmentation accuracy and architectural modifications in deep learning base models several limitations and Research challenges continue to exist. The current research focuses on Segmentation tasks alone with limited emphasis on integrating Classification mechanisms for disease diagnosis. Also, the multiple architectural advancements have been implemented individually rather than collectively. Limited research has focused on effectively integrated these complementary mechanisms within a single unified architecture.

Another significant issue still to be addressed is high computational complexity longer training duration and increased computational resources and extensive memory requirements. Furthermore, generalization capability remains an unresolved issue amongst various datasets.

Proposed work

The proposed model focuses on the combined Segmentation and Classification approach which also incorporates Attention mechanisms and Residual connections for improved efficiency and Segmentation sensitivity. The model also produces computational overhead and promotes generalization among different datasets. The proposed model will also incorporate Explainable AI (XAI) techniques to enhance the transparency and interpretability of the decision-making process in order to improve trust in automated disease diagnosis.

5. CONCLUSION AND FUTURE SCOPE

Deep Learning models continue to transform the field of Medical image analysis by enabling more efficient and accurate disease detection and diagnosis. This study provides a strong foundation for future research focused on developing generalized, clinically applicable and computationally efficient frameworks for Medical image Segmentation and Classification. Such advancements have the potential for an accurate treatment planning and improved diagnostic accuracy to improve overall healthcare outcomes.

7. RESEARCH INTEGRITY STATEMENT

The authors declare that the manuscript is original, complies with the journal's plagiarism and AI usage requirements, has not been published or submitted elsewhere, and that all applicable ethical guidelines have been followed in conducting, reporting, and publishing the research.

8. REFERENCES

[1] Ronneberger O., Fischer P., and Brox T., "U-Net: Convolutional Networks for Biomedical Image Segmentation," MICCAI, 2015.

[2] Badrinarayanan V., Kendall A., and Cipolla R., "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2017.

[3] Zhou Z., Siddiquee M.M.R., Tajbakhsh N., and Liang J., "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," Deep Learning in Medical Image Analysis (DLMIA) Workshop, 2018.

[4] Zhang Z., Liu Q., and Wang Y., "Road Extraction by Deep Residual U-Net," IEEE Geoscience and Remote Sensing Letters, 2018.

[5] Isensee F., Jaeger P.F., Kohl S.A.A., Petersen J., and Maier-Hein K.H., "nnU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation," Nature Methods, 2021.

[6] Oktay O., Schlemper J., Folgoc L.L., Lee M., Heinrich M., Misawa K., Mori K., McDonagh S., Hammerla N.Y., Kainz B., Glocker B., and Rueckert D., "Attention U-Net: Learning Where to Look for the Pancreas," Medical Imaging with Deep Learning (MIDL), 2018.

[7] Chen J., Lu Y., Yu Q., Luo X., Adeli E., Wang Y., Lu L., Yuille A.L., and Zhou Y., "TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation," arXiv, 2021.

[8] Milletari F., Navab N., and Ahmadi S.A., "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," Fourth International Conference on 3D Vision (3DV), 2016.

[9] Çiçek Ö., Abdulkadir A., Lienkamp S.S., Brox T., and Ronneberger O., "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," MICCAI, 2016.