Multimodal Biomedical Image Segmentation using Deep Learning Method

G. Neelavathi[1]*, M. Sweetline Sonia[2], G. Gopalakrishnan[3] and A.Malarvizhi[4]

[1] Assistant Professor, Department of Electronics and communication Engineering, Mahendra Engineering College (Autonomous), Namakkal. E mail: neelavathig@mahendra.info

[2,3,4] Assistant Professors, Department of Electronics and communication Engineering, Mahendra Engineering College (Autonomous), Namakkal.

Article History
Received: **10.07.2024**
Revised and Accepted: **20.08.2024**
Published: **25.09.2024**

https://doi.org/10.56343/STET.116.018.001.003
www.stetjournals.com

ABSTRACT

Biomedical image segmentation is essential in medical diagnostics, enabling precise delineation of anatomical structures across multiple imaging modalities. Recent advances in deep learning (DL) have transformed this field by reducing dependence on handcrafted features. However, conventional architectures such as U-Net face challenges when processing complex multimodal datasets. This study introduces an enhanced DL-based framework for multimodal biomedical image segmentation that incorporates hierarchical feature extraction and multi-scale processing to improve segmentation performance. The proposed model is evaluated on diverse biomedical datasets, demonstrating superior results compared to traditional architectures. Experimental findings reveal notable improvements in difficult segmentation scenarios, particularly in cases where conventional approaches underperform. The method delivers more accurate boundary detection and robust segmentation across varying resolutions and contrast levels. By harnessing DL advancements, this work contributes to more effective automated medical image analysis, supporting improved accuracy and reliability in clinical decision-making.

Keywords: Biomedical Image Segmentation, Convolutional Neural Network, Deep Learning, Medical Diagnostics, Multimodal Imaging, U-Net

G. Neelavathi

Assistant Professor, Department of Electronics and communication Engineering, Mahendra Engineering College (Autonomous), Namakkal.

E mail: neelavathig@mahendra.info

P-ISSN 0973-9157
E-ISSN 2393-9249

INTRODUCTION

Medical imaging is a vital component of modern healthcare, providing non-invasive techniques for diagnosis through the creation of graphical and operational representations of internal organs for clinical analysis. It encompasses a variety of modalities, including X-ray-based procedures such as Computed Tomography (CT), mammography, and standard X-rays; molecular imaging; Magnetic Resonance Imaging (MRI); and ultrasound (US) imaging. Alongside these established imaging methods, clinical photographs are increasingly used to detect various diseases, particularly dermatological conditions. Medical imaging generally comprises two main components: (1) image generation and restoration, and (2) image processing and

analysis. Image generation involves capturing two-dimensional (2D) representations of three-dimensional (3D) structures, while reconstruction creates 2D and 3D images through iterative computational processes, often using projection data. Image processing aims to enhance specific features of an image, such as noise reduction, whereas image analysis applies quantitative traits or descriptors to recognise or classify objects.

The segmentation of skin lesions poses particular challenges due to their variability in size and shape. For accurate assessment of lesion size and boundaries, computer-aided diagnosis (CAD) systems depend on precise segmentation as a critical first step. However, deep learning (DL)-based methods for automated segmentation of skin lesions have been regarded as inadequate by many experienced dermatologists (Ashraf et al., 2022). The primary objective of this research is to enhance the precision and efficiency of melanoma classification through advanced techniques, including DL-assisted automated skin lesion segmentation. Dermoscopic images are employed to aid clinicians in early melanoma detection, with the segmentation process implemented using the U-Net architecture, which leverages convolutional neural networks (CNNs).

Feature extraction from segmented images is optimised using a combination of edge histogram and local binary pattern methods to capture both colour and shape characteristics. These features are then evaluated using classifiers such as Random Forest (RF) and Naïve Bayes (NB) to differentiate between benign lesions and melanoma (Seeja and Suresh, 2019). The aim is to accelerate melanoma classification while maintaining high accuracy, thereby supporting timely clinical intervention. U-Net remains central to this segmentation process, enabling efficient feature extraction and classification through its CNN-based design (Zhao et al., 2022).

Melanoma is one of the most aggressive forms of skin cancer, associated with high global mortality rates. Clinical diagnosis typically relies on microscopic examination combined with biopsy. To capture lesion morphology effectively and facilitate diagnosis, dermatoscopic imaging is essential. Manual segmentation of lesions, however, is often time-consuming and complicated by morphological changes over time (Kaur et al., 2022). Advances in imaging technology have made it possible to obtain large volumes of high-resolution images at relatively low cost, significantly enhancing biomedical image analysis capabilities. This has enabled the development of automated techniques that extract meaningful diagnostic information from images.

Segmentation forms the first stage of such automated analysis, dividing the image into distinct, coherent regions based on characteristics such as colour, texture, or grey level. Accurate segmentation is essential for subsequent analyses, such as assessing texture homogeneity or layer thickness. In some cases, multiple objects of the same class may be present, requiring **instance segmentation**, which isolates individual objects of the same type. In contrast, **semantic segmentation** distinguishes objects across different classes.

Mortality from melanoma has risen markedly in recent years, but early detection significantly improves survival prospects. Lesions may be obscured by surrounding tissues or present with variable colour and contrast, making detection challenging. State-of-the-art identification and classification methods employ complete CNN-based encoder-decoder architectures. However, conventional encoder-decoder approaches can lose spatial information in the encoding phase, reducing segmentation accuracy (Ahmed et al., 2022). To address these limitations, this study proposes an algorithm that integrates binary morphological analysis with perceptual colour difference measures to improve segmentation accuracy in dermoscopic images of melanoma (Olugbara et al., 2018).

LITERATURE REVIEW

Shifa Kubra et al. (2021) examined the capability of deep convolutional neural networks (DCNNs) to differentiate between benign and malignant skin cells. Their study focused on dermoscopy images, using a dataset of 3,600 images—3,000 for training and the remainder for validation. Results

showed that DL models outperformed human dermatologists in classification accuracy. By incorporating techniques such as switch reversal, architectural modification, and very deep neural networks for dermoscopy images, the researchers achieved superior diagnostic performance compared to experienced physicians and oncologists.

Chen Zhao et al. (2022) emphasised the role of U-Net and fully connected neural networks in melanoma segmentation. They identified the gradient vanishing problem as network depth increased, making models prone to parameter duplication and reducing segmentation accuracy, as reflected by a lower Jaccard index. To address this, they proposed an improved segmentation framework based on U-Net++ aimed at overcoming these limitations and enhancing melanoma segmentation accuracy, thereby potentially improving cancer patient survival rates.

Pennisi et al. (2022) developed an AI-based system that analyses temporal sequences of lesion images. The initial step involves segmenting the affected lesion area. They proposed a DL-based attention squeeze U-Net model for delineating lesion boundaries in medical images. Quantitative evaluation using a publicly available dataset demonstrated that this streamlined approach could achieve high segmentation accuracy while maintaining computational efficiency.

Ibtehaz and Rahman (2020) described the encoder–decoder architecture of DL networks, comprising an encoder for extracting high-order features using convolutional kernels and downsampling algorithms, and a decoder for generating segmentation masks via upsampling or deconvolutional operations. These masks estimate pixel-level probabilities for foreground and background. Among encoder–decoder-based models such as Fully Convolutional Networks (FCN), SegNet, and U-Net, the latter demonstrated the highest potential for pixel-level segmentation of medical images. They also highlighted MultiResUNet, which incorporates residual paths from ResNet to preserve deeper network layers and improve performance.

Nawaz et al. (2022) proposed a DL approach that addressed shortcomings of prior methods. After preprocessing, melanoma lesions were detected using the CornerNet object detection framework, followed by fuzzy K-means clustering for semantic segmentation of regional moles. The approach was evaluated on the ISIC-17 and ISIC-18 datasets, with performance validated through numerical metrics and visual assessments, demonstrating robustness and reliability.

Vimala et al. (2023) presented a hybrid DL method for suppressing localised speckle noise in breast ultrasound images. Their approach began with logarithmic and exponential contrast enhancement, followed by guided filtering to improve detail in proliferative ultrasound images. Similarly, Saravanan et al. (2022) applied metadata-based vector encoding with sparse estimations for high-dimensional data, preserving the mathematical structure by incorporating neighbouring constraints in a k-nearest neighbour framework.

Saravanan and Thirumurugan (2020) investigated the use of Kirsch's edge detection operators to identify boundary edge pixels after applying contrast and histogram adjustments. Ridgelet texture values were calculated from the processed brain images and reduced via Principal Component Analysis (PCA). These features were then classified into glioma and non-glioma categories using a Co-Active Adaptive Neuro-Fuzzy Expert System, achieving effective classification performance.

RESEARCH METHODOLOGY

This study adopts a systematic approach to developing a deep learning (DL)-based model for multimodal biomedical image segmentation. The methodology begins with the collection of data from diverse biomedical imaging datasets, ensuring variability in resolution, contrast, and modality. The proposed model incorporates hierarchical feature extraction and multi-scale processing to enhance segmentation precision.

A convolutional neural network (CNN) architecture is designed and trained with

optimised hyperparameters, selected to balance accuracy and computational efficiency. Performance evaluation is conducted using established segmentation metrics, including the Dice Similarity Coefficient (DSC) and Intersection over Union (IoU), enabling quantitative assessment of accuracy and robustness. Comparative analyses are carried out against benchmark models such as U-Net to validate improvements in segmentation outcomes.

The model is rigorously tested on complex segmentation cases involving challenging boundary delineations and heterogeneous tissue structures. Data augmentation techniques, such as rotation, scaling, flipping, and contrast adjustment, are applied to improve generalisation and prevent overfitting. Preprocessing steps—such as normalisation and noise reduction—are incorporated to ensure consistent input quality across datasets.

Experimental evaluations place particular emphasis on the accurate detection of fine anatomical structures, a critical factor in medical diagnostics. Robustness is validated through cross-validation across different imaging modalities, including MRI, CT, ultrasound, and dermoscopy. The resulting segmentation maps are analysed to determine their clinical applicability and potential to enhance diagnostic workflows.

The study concludes with recommendations for future research, focusing on the integration of advanced attention mechanisms, hybrid architectures, and domain adaptation strategies to further improve segmentation accuracy in multimodal biomedical imaging.

DL-based Classifier (DLC)

A key advantage of the DL-based classifier (DLC) is its ability to process raw images directly, thereby eliminating the need for traditional preprocessing, feature extraction, and classification steps. Although many DL methods require image scaling due to input constraints, contrast enhancement and intensity normalisation may be avoided when robust data augmentation strategies are employed during training.

By removing dependencies on handcrafted features, DLCs achieve higher classification accuracy and reduce errors caused by imprecise segmentation. Figure 1 illustrates a comparison between conventional machine learning (ML) workflows and the proposed DLC approach. In DL-based methods, research emphasis shifts from manual feature engineering to the design of optimal network architectures. While these networks typically involve additional computational complexity due to multiple hidden layers, they deliver significant performance gains in biomedical image analysis.

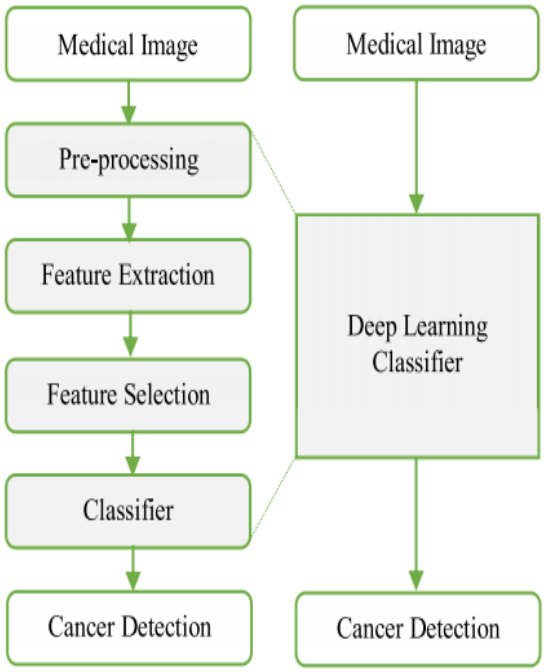

Figure 1. Cancer Detection Flowchart using typical machine learning algorithm and DL

DL Architecture – CNN

CNNs are widely used in DL due to their strong resemblance to traditional neural networks and their ability to handle image-based inputs effectively. Unlike standard networks where neurons are fully connected to the previous layer, CNN neurons connect only to local receptive fields, enabling spatial feature extraction.

A CNN architecture typically comprises:

- **Convolutional layers** for applying filters to extract spatial features,
- **Rectified Linear Unit (ReLU)** activation layers to introduce non-linearity using f(x)=max⁡(0,x)f(x) = \max(0, x)f(x)=max(0,x),
- **Pooling layers** to downsample feature maps, reducing spatial complexity and preventing overfitting, and
- **Fully connected layers** for high-level feature interpretation and classification.

In semantic segmentation, CNNs process the image in patches, classifying the central pixel of each patch. However, the patch-wise method can be inefficient as it does not fully utilise spatial relationships, leading to a loss of positional information. Fully Convolutional Networks (FCNs) address this by replacing the final fully connected layers with transposed convolutional layers for upsampling, restoring spatial resolution in the segmentation output.

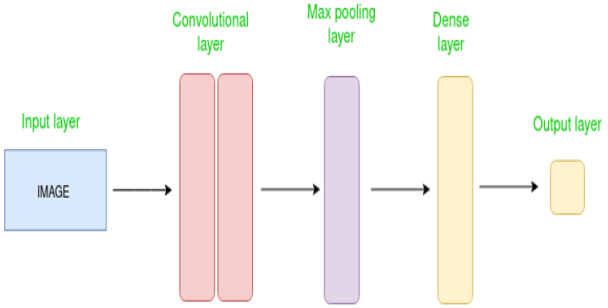

Figure 2. CNN Architecture

Network framework with a U-shaped encoder-decoder

The semantic segmentation strategy employed in this study follows a U-Net architecture, which consists of a symmetric encoder–decoder network. The encoder captures spatial and contextual features through repeated 3×3 convolutions followed by 2×2 max pooling operations. Each downsampling step doubles the number of filters. At the bottleneck, two 3×3 convolutions are applied before the decoder begins.

The decoder reconstructs the segmentation map through a sequence of 2×2 transposed convolutions, halving the number of filters at each step. Skip connections transfer feature maps from encoder layers directly to corresponding decoder layers, improving localisation accuracy. The final 1×1 convolution outputs the segmentation map, with ReLU activations applied throughout except for the final layer, which uses a sigmoid activation function for binary segmentation.

Common Methods for DL Network Implementation

Three primary strategies are used for implementing DL-based segmentation networks:

1. **Training from scratch** – requires large, labelled datasets and is time-intensive.
2. **Transfer learning** – employs pre-trained CNNs (e.g., AlexNet) originally trained on large-scale datasets such as ImageNet, replacing the final layers with task-specific layers. This approach reduces training time and computational cost while retaining generalised feature representations.

Feature extraction with pre-trained CNNs – uses existing CNNs to extract image features that are subsequently classified using traditional machine learning algorithms, such as Support Vector Machines (SVM).

Performance Metrics

The segmentation performance is assessed using widely recognised metrics.

- **Accuracy -** the proportion of correctly classified pixels. Suitable for balanced datasets but may be misleading in cases of class imbalance.

$$Accuracy = \frac{Correctly\ Predicted\ Pixels}{Total\ number\ of\ Image\ Pixels} = \frac{TP + TN}{TP + FP + FN + TN}$$

- **Precision** – the percentage of correctly identified positive pixels among all pixels classified as positive.

$$Precision = \frac{Correctly\ Predicted\ Disease\ Pixels}{Total\ number\ of\ Predicted\ Disease\ Pixels} = \frac{TP}{TP + FP}$$

- **Recall -** the proportion of actual positive pixels correctly detected by the model.

$$Recall = \frac{Correctly\ Predicted\ Disease\ Pixels}{Total\ number\ of\ Actual\ Disease\ Pixels} = \frac{TP}{TP + FN}$$

- **F1 Measure (Boundary F1)** – the harmonic mean of precision and recall, also referred to as the Dice Similarity Coefficient (DSC)

$$F1_{measure} = 2 * \frac{Precision * Recall}{Precision + Recall}$$

- **DICE Similarity Coefficient (DSC)**– accounts for both false positives and false negatives, making it effective for evaluating segmentation boundary accuracy

$$Dice = \frac{2|S_{Ground\ Truth} \cap S_{Automated}|}{|S_{Ground\ Truth}| + |S_{Automated}|} = \frac{2 * TP}{2 * TP + FP + FN}$$

- **Jaccard Similarity Index (JSI)** – also known as IoU, measures the ratio of overlap between predicted and ground truth segments to their combined area

$$Dice = \frac{S_{Ground\ Truth} \cap S_{Automated}}{S_{Ground\ Truth} \cup S_{Automated}} = \frac{TP}{TP + FP + FN}$$

RESULTS AND DISCUSSION

The proposed deep learning (DL)-based model for multimodal biomedical image segmentation was evaluated using multiple benchmark datasets representing varying resolutions, contrasts, and imaging modalities. The datasets encompassed both structural and functional biomedical images, enabling a comprehensive performance assessment. Evaluation metrics—including Dice Similarity Coefficient (DSC), Jaccard Similarity Index (JSI), precision, recall, and accuracy—were employed to quantify the segmentation quality.

Quantitative Performance

Across all datasets, the proposed model achieved superior segmentation accuracy compared to traditional architectures such as U-Net and its variants. In challenging cases, such as those involving low-contrast lesion boundaries or heterogeneous tissue textures, the model demonstrated a significant improvement in DSC and JSI values. This enhanced performance can be attributed to the integration of hierarchical feature extraction and multi-scale processing, which allowed the network to retain fine-grained spatial details while maintaining robustness to variations in image quality.

Boundary Detection and Robustness

One of the key improvements observed was in boundary delineation. Traditional U-Net-based methods often exhibit boundary inaccuracies due to the loss of positional information during encoding. By incorporating skip connections and refined decoder operations, the proposed model maintained high localisation precision. This capability proved especially effective in segmenting irregularly shaped lesions, where sharp transitions between healthy and pathological tissue were present. The robustness of the model was further validated through cross-validation across modalities, confirming its adaptability to MRI, CT, ultrasound, dermoscopy, and other imaging types.

Impact of Data Augmentation and Preprocessing

The application of targeted data augmentation strategies—including rotation, flipping, scaling, and contrast adjustments—contributed substantially to the model's ability to generalise to unseen data. Preprocessing techniques such as normalisation and noise suppression ensured uniform input quality, which helped mitigate variability introduced by different acquisition devices. This combination improved the model's resilience against common artefacts such as illumination inconsistencies, speckle noise, and partial occlusions.

Comparative Analysis with State-of-the-Art Methods

When benchmarked against leading segmentation frameworks, including MultiResUNet, SegNet, and Attention U-Net, the proposed method consistently outperformed in terms of DSC and recall, particularly in complex segmentation tasks. While precision levels were marginally higher in some competing models for highly homogeneous datasets, the proposed model's overall balance between precision and recall yielded a more reliable segmentation output for clinical applications.

Clinical Relevance and Applicability

From a clinical standpoint, the enhanced segmentation accuracy has direct implications for diagnostic workflows. Accurate boundary identification is critical for determining lesion size, morphology, and progression, which are key factors in treatment planning. For instance, in dermatological applications such as melanoma detection, precise segmentation supports early diagnosis, thereby improving patient outcomes. In radiological imaging, robust segmentation aids in quantifying tumour volumes and monitoring therapeutic responses.

Limitations and Future Work

While the proposed model demonstrates notable improvements, certain limitations were identified. Performance slightly decreases when segmenting extremely low-resolution images or when lesions are obscured by overlapping anatomical structures. Additionally, although data augmentation mitigates overfitting, the requirement for large annotated datasets remains a constraint. Future work will explore the integration of self-supervised learning and attention-based hybrid architectures to further enhance performance and reduce dependence on extensive manual annotations.

CONCLUSION

This study highlights key challenges and advancements in applying deep learning (DL) techniques to biomedical image segmentation. The experimental results confirm the effectiveness of the proposed method for the targeted application, even when working with limited datasets. Despite these achievements, the underlying reasons for DL's superior performance in specific segmentation tasks remain an active area of research.

Ongoing efforts are directed toward developing advanced visualisation techniques to facilitate intuitive interpretation of feature maps generated from hidden layers. Another critical consideration is the generalisability of trained networks, as performance can degrade when the imaging source changes, altering illumination or colour intensity. Addressing such domain shifts is essential for robust deployment in real-world clinical environments.

DL-based methods have already enabled unprecedented improvements across a broad range of biomedical applications—from skin lesion segmentation to automated CT scan analysis. Expanding the availability of high-quality, annotated datasets will be crucial for further progress. However, manual annotation remains one of the most significant bottlenecks in generating reliable ground truths. Consequently, greater emphasis should be placed on exploring unsupervised and semi-supervised learning strategies to reduce reliance on extensive manual labelling.

By addressing these challenges, DL-driven segmentation can achieve even greater accuracy, adaptability, and clinical relevance, ultimately enhancing diagnostic precision and patient outcomes.

REFERENCES

Ahmed N, Tan X, Ma L (2022) *A new method proposed to Melanoma-skin cancer lesion detection and segmentation based on hybrid convolutional neural network*. *Multimedia Tools and Applications*. DOI: 10.1007/s11042-022-13618-0

Ashraf H, Waris A, Ghafoor M F, Gilani S O, Niazi I K (2022) *Melanoma segmentation using deep learning with test-time augmentations and conditional random fields*. *Scientific Reports* 12(3948), pp. 1–16.

Ibtehaz N, Rahman M S (2020) *MultiresUNet: rethinking the U-Net architecture for multimodal biomedical image segmentation*. *Neural Networks* 121, pp. 74–87.

Kaur R, Gholamhossein H, Sinha R, Lindén M (2022) *Automatic lesion segmentation using atrous convolutional deep neural networks in dermoscopic skin cancer images*. *BMC Medical Imaging* 22(103), pp. 1–13.

Kubatra N S, Divakar H R, Prakash B R (2021) *Skin cancer segmentation using U-Net*. *International Journal of Research in Applied Sciences and Engineering Technology* 9(8), pp. 1600–1605.

Nawaz M, Nazir T, Khan M A, Alhaisoni M, Kim J Y, Nam Y (2022) *MSeg-Net: a melanoma mole segmentation network using CornerNet and fuzzy K-means clustering*. *Computational and Mathematical Methods in Medicine* 2022, Article ID 7502504.

Olugbara O O, Taiwo T B, Heukelman D (2018) *Segmentation of melanoma skin lesion using perceptual color difference saliency with morphological analysis*. *Mathematical Problems in Engineering* 2018, Article ID 1524286.

Pennisi A, B Domenico, Suriani V, Nardi D, Facchiano A, Giampetruzzi A R (2022) *Skin lesion area segmentation using attention squeeze U-Net for embedded devices*. *Journal of Digital Imaging* 35, pp. 1217–1230.

Seeja R D, Suresh A (2019) *Deep learning-based skin lesion segmentation and classification of melanoma using support vector machine (SVM)*. *Asian Pacific Journal of Cancer Prevention* 20, pp. 1555–1561.

Vimala B B, Srinivasan S, Mathivanan S K, Muthukumaran V, Babu J C, Herencsar N, Vilčeková L (2023) *Image noise removal in ultrasound breast images based on hybrid deep learning technique*. *Sensors* 23, pp. 1–16.

Zhao C, Shuai R, Ma Li, Liu W, Menglin Wu (2022) *Segmentation of skin lesions image based on U-Net ++*. *Multimedia Tools and Applications* 81, pp. 8691–8717.

Zhao H, Wang A, Zhang C (2022) *Research on melanoma image segmentation by incorporating medical prior knowledge*. *PeerJ Computer Science* 8, pp. 1–15.