Transformer-Enhanced Multi-Scale Learning: A Hybrid Convolutional-Transformer Architecture for Robust Wound Segmentation

Prathamesh Gadekar
Department of Computer Science and Engineering
Indian Institute of Information Technology, Pune
Pune, India
112115115@cse.iiitp.ac.in

Sanjeev Sharma
Associate Dean
Indian Institute of Information Technology, Pune
Pune, India
sanjeevsharma@iiitp.ac.in

***Abstract*—Accurate wound segmentation is critical for monitoring chronic conditions like diabetic foot ulcers, where precise area measurement directly impacts treatment planning and healing assessment. While deep learning has advanced medical image analysis, existing wound segmentation methods face challenges in handling diverse wound morphologies and small lesion sensitivity. We propose WoundSegFormerV2, a novel architecture integrating ConvNeXt-based feature extraction, Dynamic Atrous Spatial Pyramid Pooling (DASPPV2), and global transformer context blocks with attention-based refinement. Key innovations include learnable branch weights in DASPPV2 for adaptive multi-scale context fusion, CBAM-enhanced decoder blocks for boundary precision, and an ensemble strategy leveraging multiple backbone configurations. Evaluated on existing AZH dataset (containing 1,109 foot ulcer images) with rigorous clinical annotations, our model achieves state-of-the-art performance with a Dice score of 92.97% and IoU of 86.95%, outperforming prior U-Net variants and hybrid architectures. The proposed framework demonstrates robustness to varying lighting conditions and irregular wound boundaries while maintaining computational efficiency. These advancements enable reliable integration into clinical workflows, offering potential applications in automated wound documentation. By addressing critical gaps in generalizability and precision, this work represents a significant step toward standardized, data-driven wound care.**

***Index Terms*—Deep-Learning, wound-segmentation, healthcare, medical-image-segmentation, wound-analysis**

I. Introduction

Chronic wounds affect over 10.5 million people globally. Traditional manual segmentation by clinicians is time-consuming and subjective, driving demand for automated solutions [1]. Wound image segmentation – the pixel-wise delineation of wound regions from surrounding tissue – serves as the foundation for quantitative analysis of healing progress, infection detection, and treatment personalization. Despite progress in medical imaging, wound segmentation remains uniquely challenging due to irregular morphologies, variable lighting conditions, and low contrast between healthy and damaged tissue.

In the United States alone, chronic wounds burden more than 10.5 million Medicare beneficiaries—an increase of 2.3 million since 2014—and now impact nearly 2.5% of the entire population [1]. This proportion climbs even higher among older adults, driven by the increasing prevalence of age-related comorbidities such as diabetes and obesity. Chronic wounds not only diminish quality of life but also contribute significantly to healthcare costs due to their tendency to cause complications, prolong hospital stays, and require intensive management. With global trends pointing toward an aging population and a persistent threat from metabolic and infectious diseases, the clinical, social, and economic burden of chronic wounds is expected to escalate further.

Despite the critical importance of reliable wound-specific segmentation for objective monitoring and treatment planning, the vast majority of deep-learning research has centered on broader medical image segmentation tasks—such as organ and tumor delineation [2] [3], brain-tumor segmentation [4], cell nuclei segmentation [5], lung parenchyma segmentation [6], and pulmonary nodule analysis [7]—leaving wound images relatively under-studied. Recent advances in automated wound segmentation leverage enhanced U-Net architectures—often with transfer learning, ensembles, or hybrid encoders to achieve impressive results, but struggle with small lesions and noisy labels. Efforts to improve generalizability include new datasets (WoundSeg) [8], adaptive pretraining, and global–local feature fusion. Widely used datasets such as Medetec [9] are limited in size and diversity, leading to biases and robustness issues.

This work addresses two core challenges: (1) improving segmentation accuracy across diverse wound presentations, and (2) enhancing model robustness to dataset noise. We introduce WoundSegFormerV2, a hybrid architecture that synergizes ConvNeXt's [10] hierarchical feature learning with transformer-based global context modeling [11]. Our key innovations include dynamic atrous spatial pyramid pooling [12], global transformer context blocks, and CBAM-enhanced decoder pathways [13], supported by ensemble learning across multiple ConvNeXt backbones.

The proposed model achieves a Dice score of 92.97% and IoU of 86.95% on AZH dataset [14], surpassing previous state-of-the-art methods.

II. Related Work

Automated wound segmentation is critical for managing chronic conditions like diabetes, where manual clinical assessments are error-prone [15]. Recent studies have advanced deep learning approaches to address challenges in dataset diversity, model generalizability, and computational efficiency. A prevalent strategy involves enhancing U-Net architectures with transfer learning, data augmentation, and ensemble techniques. For instance, ensemble frameworks combining LinkNet and U-Net with EfficientNet backbones achieved impressive Dice scores (Dice: 92.09%) on chronic wound datasets [16] but struggled with small lesions and dataset noise. Hybrid models like M-Vgg19-Unet integrated hierarchical feature extraction (VGG19) with U-Net's encoder-decoder structure, attaining superior accuracy (Dice: 92.02%) through parameter-rich designs, albeit at higher computational costs [17].

Efforts to improve generalizability include novel datasets such as WoundSeg (2,686 images across eight wound types) and architectures like WSNet, which employs adaptive pretraining and global-local feature fusion [8]. These innovations yielded robust performance (Dice: 84.7%) across diverse wound morphologies and imaging conditions. Lightweight solutions, such as MobileNetV2 with connected component labeling [14], balanced accuracy (Dice: 90.47%), enabling mobile deployment. Attention mechanisms further refined segmentation; Dual Attention VGG16-U-Net combined spatial and channel-wise recalibration to achieve 94.1% Dice on WTS Dataset [18].

Dataset Limitations remain a critical barrier. While several studies have employed deep learning for ulcer segmentation and classification some rely on minimal corpora. For example, the Medetec wound dataset [9], widely used in prior work [19], [20], contains only 607 images, resulting in systems prone to brittleness. Broader limitations persist across studies: missing small lesions [16]. While augmentation mitigate data scarcity, broader validation across ethnicities and wound subtypes remains critical [16]. Collectively, these works underscore the potential of deep learning to standardize wound care.

III. Dataset

A. AZH Wound Dataset

The AZH Wound Dataset [14] comprises 1,109 foot ulcer images collected over two years from 889 patients at the Advancing the Zenith of Healthcare (AZH) Wound and Vascular Center. Images were captured using Canon SX 620 HS cameras and iPad Pro devices under uncontrolled lighting conditions. Wound regions were localized using a YOLOv3-based bounding box model and cropped to 224×224 patches with zero-padding. Segmentation masks were manually annotated by wound care specialists and rigorously validated to ensure clinical accuracy.

B. Preprocessing and Augmentation

Images and masks were resized to 256×256 pixels during training and validation. The following transformations were applied to the training set to improve model robustness:

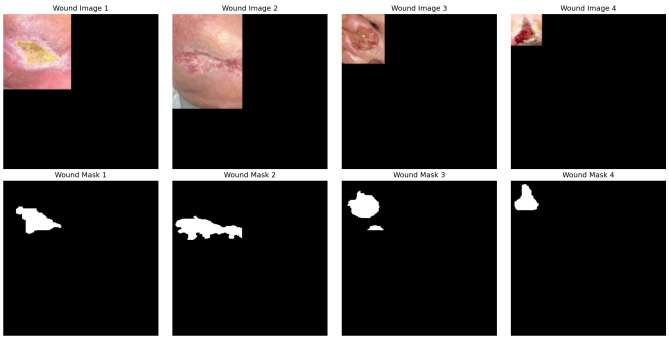

Fig. 1: Sample foot ulcer images from the AZH Wound Dataset.

- Random horizontal/vertical flipping (50% probability),
- Rotation within $\pm 30°$,
- Color jitter (brightness, contrast, saturation, and hue adjustments),
- Normalization using ImageNet mean and standard deviation.

Validation data underwent only resizing and normalization.

Furthermore, the dataset was partitioned into training and testing sets using a 75%:25% split to evaluate model performance on unseen data.

IV. Model Architecture

Our proposed framework, WoundSegFormerV22, leverages advanced neural network components to effectively segment wound regions in clinical images. The architecture consists of four primary components: (1) a ConvNeXt-based feature extraction backbone, (2) an improved Dynamic Atrous Spatial Pyramid Pooling (DASPPV2) module, (3) a Global Transformer Context Block, and (4) an enhanced decoder pathway with CBAM-based refinement. We implement ensemble learning by utilizing multiple backbone variations to maximize segmentation performance.

A. Feature Extraction with ConvNeXt Backbones

To extract multi-scale features from wound images, we employ the ConvNeXt architecture [10], a state-of-the-art CNN model that demonstrates superior performance on vision tasks. ConvNeXt modernizes traditional convolutional networks with design principles adapted from Vision Transformers while maintaining the inductive biases of CNNs.

We implement four variations of our model using different ConvNeXt backbones:

- ConvNeXt-Tiny (96-192-384-768 channels)
- ConvNeXt-Small (96-192-384-768 channels)
- ConvNeXt-Base (128-256-512-1024 channels)
- ConvNeXt-Large (192-384-768-1536 channels)

Each backbone extracts hierarchical features at four scales with progressive downsampling (1/4, 1/8, 1/16, and 1/32 of the input resolution). For a 256×256 input image, these feature maps have spatial dimensions of 64×64, 32×32, 16×16, and 8×8, respectively.

B. Dynamic Atrous Spatial Pyramid Pooling with Learnable Branch Weights (DASPPV2)

To effectively capture multi-scale contextual information in wound regions, we introduce an improved Dynamic Atrous Spatial Pyramid Pooling module (DASPPV2). This module applies parallel dilated convolutions with different dilation rates (1, 6, 12, 18, and 24) to the deepest feature map, enabling the network to incorporate context at multiple receptive fields.

DASPPV2 extends the traditional ASPP approach [12] with three key innovations:

1) **Depthwise Separable Convolutions**: Each branch first applies a depthwise convolution with the specified dilation rate, followed by a pointwise (1×1) convolution, reducing computational complexity while maintaining performance.
2) **Learnable Branch Weights**: We incorporate learnable scalar weights for each branch, allowing the network to adaptively emphasize different receptive fields based on their importance for wound segmentation. These weights are normalized using softmax to ensure proper scaling.
3) **Global Pooling Context**: An additional branch performs global average pooling followed by a 1×1 convolution, capturing global context information that is then upsampled and combined with other branches.

The outputs from all branches are weighted, concatenated, and fused through a 1×1 convolution, producing a feature map that integrates multi-scale contextual information.

C. Global Transformer Context Block

To enhance long-range dependencies and global context modeling, we incorporate a Global Transformer Context Block following the DASPPV2 module. This transformer-based component processes spatial features with self-attention mechanisms, allowing the model to establish relationships between distant spatial locations in the image.

The block operates by:

1) Reshaping the feature map into a sequence of tokens
2) Applying layer normalization
3) Processing tokens with multi-head self-attention
4) Adding a residual connection
5) Further processing with a feed-forward network (FFN) and another residual connection

This design integrates the global reasoning capabilities of transformers while maintaining the spatial structure essential for accurate segmentation.

D. Enhanced Decoder with CBAM-based Refinement

Our decoder pathway employs Dense Decoder Blocks with Convolutional Block Attention Modules (CBAM) [13] to progressively recover spatial information while refining features at each resolution level. Each DenseDecoderBlockV2 performs several key operations:

1) **Feature Upsampling**: The lower-resolution features from the previous stage are upsampled by a factor of 2.

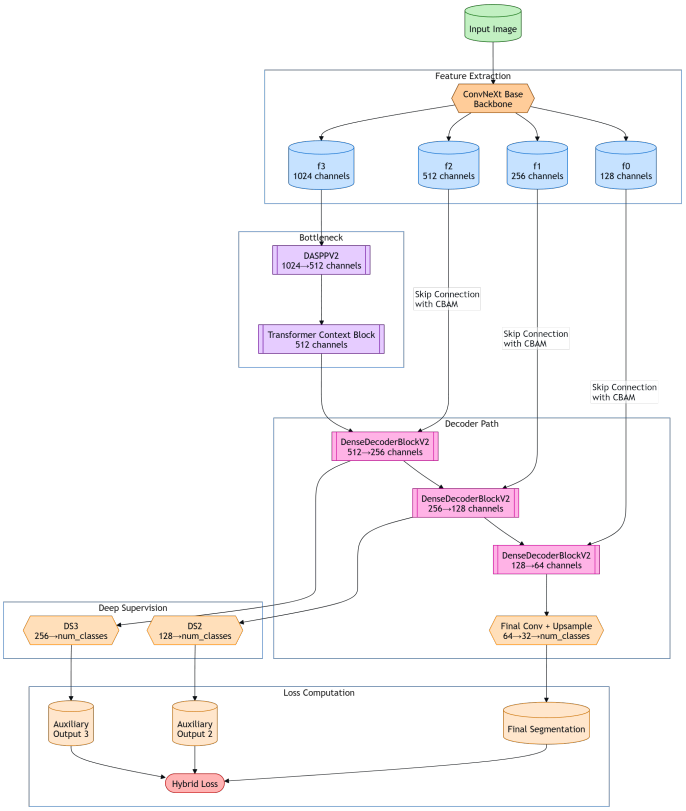

Fig. 2: Model Architecture Diagram

2) **Skip Connection Enhancement**: Features from the corresponding encoder level are refined using CBAM, which applies channel and spatial attention mechanisms sequentially.
3) **Feature Fusion**: The upsampled features are concatenated with the refined skip features.
4) **Multi-scale Feature Refinement**: Two convolutional blocks with residual connections process the fused features.
5) **CBAM Refinement**: A final CBAM module is applied to emphasize the most informative regions in the feature map.

The CBAM module consists of two sequential attention mechanisms: channel attention followed by spatial attention. The channel attention computes feature statistics using both average-pooled and max-pooled features, while spatial attention generates a spatial attention map highlighting informative regions. This design enables effective high-resolution feature recovery while maintaining the semantic information extracted by deeper layers.

E. Deep Supervision and Model Outputs

To optimize gradient flow and improve feature learning at different scales, we implement deep supervision through auxiliary segmentation heads at multiple decoder levels. The model produces three outputs:

1) **Main Output**: The final segmentation mask at the original input resolution
2) **Auxiliary Output 1**: An intermediate segmentation prediction from decoder stage 2
3) **Auxiliary Output 2**: An intermediate segmentation prediction from decoder stage 3

During training, losses are computed for all three outputs and combined with appropriate weighting factors. At inference time, only the main output is used for the final segmentation.

F. Model Ensemble Strategy

To further enhance segmentation performance, we implement an ensemble approach that combines predictions from four WoundSegFormerV2 models with different backbone configurations (ConvNeXt-Tiny, Small, Base, and Large). The ensemble strategy aggregates predictions through weighted averaging based on the validation performance of each model. This approach leverages the complementary strengths of different backbone capacities, improving the robustness of wound segmentation.

G. Evaluation Metrics

We evaluate our wound segmentation model using two standard metrics: Dice Similarity Coefficient (DSC) and Intersection over Union (IoU).

1) Dice Similarity Coefficient: DSC measures the overlap between predicted segmentation and ground truth:

$$DSC(pred, target) = \frac{2 \cdot |pred \cap target|}{|pred| + |target|} \tag{1}$$

2) Intersection over Union: IoU quantifies the ratio of overlap to the combined area:

$$IoU(pred, target) = \frac{|pred \cap target|}{|pred \cup target|} \tag{2}$$

Both metrics range from 0 to 1, with higher values indicating better performance. For implementation, we use a threshold of 0.5 on prediction probabilities and incorporate a small smoothing factor $\epsilon = 10^{-6}$ to prevent division by zero. Metrics are calculated per image and averaged across the test dataset.

H. Loss Function

The proposed model utilizes a comprehensive hybrid loss function (HybridLossV2) that combines multiple complementary components to address the challenges in wound segmentation. This loss function is designed to enhance boundary precision while maintaining overall segmentation quality through deep supervision.

The HybridLossV2 function integrates five distinct loss components:

1) **Dice Loss**: Optimizes the overlap between predicted and ground truth segmentation masks. Based on the DSC metric, it penalizes poor overlap between prediction and target.
2) **Focal Loss**: Implemented as binary cross-entropy loss to focus training on difficult-to-classify pixels, improving the model's performance on challenging wound boundaries. This loss weights misclassified pixels more heavily than correctly classified ones.
3) **Boundary Loss**: Specifically, target edge regions by applying convolution operations to extract boundary information from both prediction and ground truth masks, enhancing the precision of wound contours. It also helps to focus on small lesions.
4) **IoU Loss**: Incorporated to improve spatial consistency of segmentation results. This loss, based on the IoU metric, helps optimize for better overlap while minimizing false positives and negatives.
5) **Deep Supervision Loss**: Applied to intermediate decoder outputs (ds_2 and ds_3), encouraging meaningful feature representations at multiple scales throughout the network.

The final loss function is formulated as:

$$\mathcal{L} = \mathcal{L}_{main} + \gamma(\mathcal{L}_{ds_3} + \mathcal{L}_{ds_2}) \tag{3}$$

Where $\mathcal{L}_{main}$ combines Dice, Focal, Boundary, and IoU losses for the main prediction:

$$\mathcal{L}_{main} = \mathcal{L}_{Dice} + \mathcal{L}_{BCE} + \alpha \cdot \mathcal{L}_{Boundary} + \beta \cdot \mathcal{L}_{IoU} \tag{4}$$

Of these components, we present the detailed equations for the Boundary Loss and Deep Supervision Loss:

Boundary Loss:

$$\mathcal{L}_{Boundary} = \frac{1}{N} \sum_{i=1}^{N} |E_{pred,i} - E_{target,i}| \tag{5}$$

$$E_{pred} = edge(pred) \tag{6}$$

$$E_{target} = edge(target) \tag{7}$$

Where N is the total number of pixels and the function $edge(\cdot)$ represents the boundary extraction operation implemented using convolution with a 3×3 kernel.

Deep Supervision Loss:

$$\mathcal{L}_{ds_k} = 1 - \frac{1}{B} \sum_{i=1}^{B} \frac{2 \cdot I_{ds_k,i} + \epsilon}{U_{ds_k,i} + \epsilon} \tag{8}$$

$$I_{ds_k,i} = \sum_{j}^{H,W} pred_{ds_k,i,j} \cdot target_{i,j} \tag{9}$$

$$U_{ds_k,i} = \sum_{j}^{H,W} pred_{ds_k,i,j} + \sum_{j}^{H,W} target_{i,j} \tag{10}$$

Where B is the batch size, H and W are the height and width of the image, and ϵ is a small smoothing constant to prevent division by zero. The deep supervision loss applies a Dice loss to intermediate decoder outputs, encouraging feature learning at multiple scales.

The hyperparameters α, β, and γ control the relative contribution of boundary loss, IoU loss, and deep supervision respectively. This composite loss function guides the network

to produce accurate wound segmentations with well-defined boundaries while leveraging information across multiple resolution scales.

V. Results

A. Experimental Setup and Training Dynamics

The proposed WoundSegFormerV2 models were trained using the AdamW optimizer with a batch size of 16 and initial learning rate of 1×10^{-4}, reduced via cosine annealing. Models using ConvNeXt-Tiny, Small, and Base backbones were trained for 400 epochs, while the ConvNeXt-Large variant required fewer epochs (330) due to computational constraints while maintaining convergence stability.

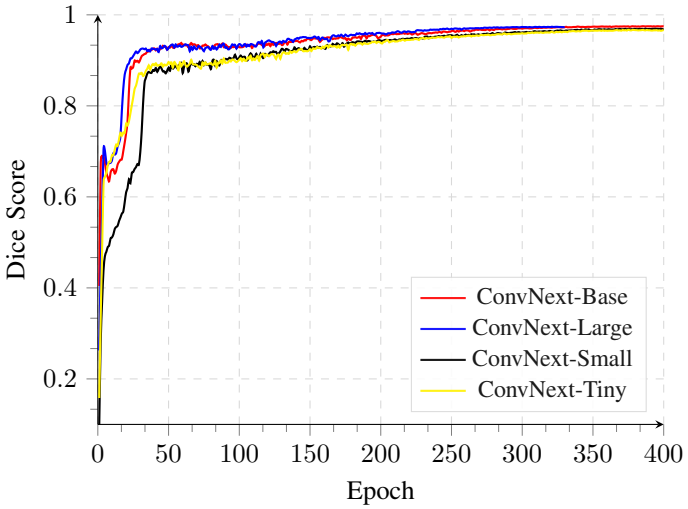

Fig. 3: Training Dice scores across different feature extractors.

Training dynamics across backbone configurations (Figures 3, 4, and 5) demonstrate stable convergence with consistent metric improvements. The Dice score progression (Figure 3) shows rapid initial gains. Loss curves (Figure 4) exhibit smooth decay patterns without significant oscillations, indicating effective regularization from the hybrid loss components. All backbones achieved steady progress throughout the training process (Figure 5), with larger models (Base/Large) showing marginally faster convergence due ti their increased parameter counts.

B. Comparative Performance Analysis

Table I compares our ensemble variants against state-of-the-art wound segmentation models. The proposed architecture achieves superior performance, with the ConvNeXt-Tiny/Base/Large ensemble attaining a Dice score of 92.97% - a 0.27% absolute improvement over the previous best WDAP LinkNet model. Notably, even smaller ensemble combinations (Base/Large) outperform all baseline approaches, demonstrating the effectiveness of our hybrid convolutional-transformer design.

The exceptional performance corresponds to precise boundary delineation across diverse wound morphologies. Comparative analysis reveals two key advantages: 1) Enhanced small

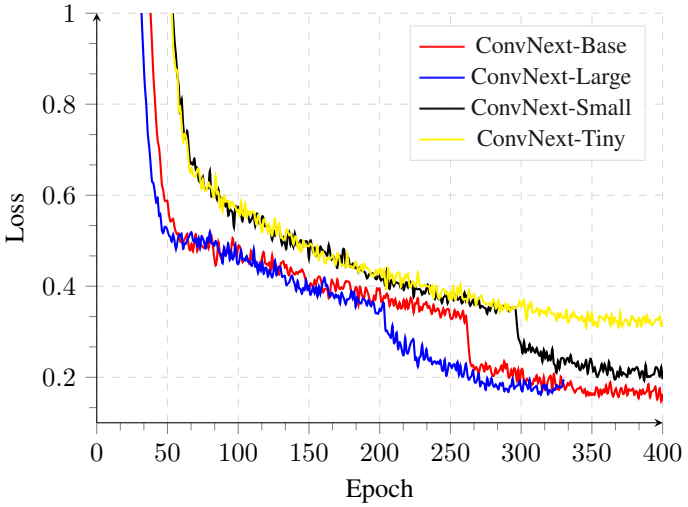

Fig. 4: Training Loss across different feature extractors.

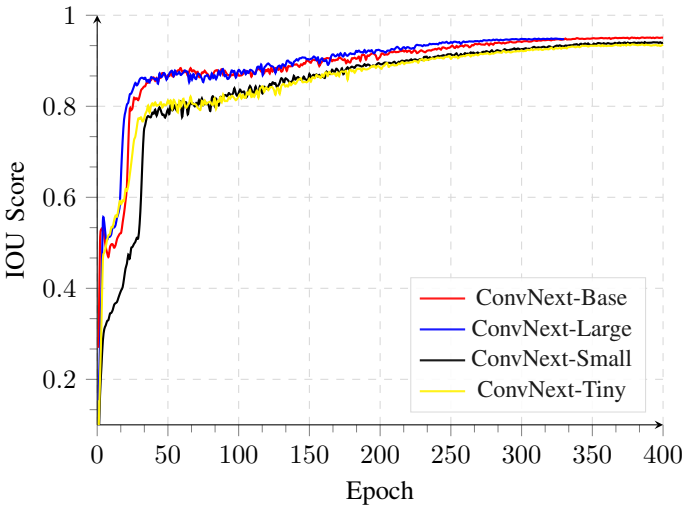

Fig. 5: Training IOU scores across different feature extractors.

lesion detection through boundary-aware loss components, and 2) Improved robustness to lighting variations via ConvNeXt's hierarchical feature learning combined with transformer-based global context modeling.

C. Ensemble Strategy Effectiveness

Our weighted ensemble strategy significantly boosts performance over individual backbones (Tiny: 91.23%, Large: 92.15%). The optimal three-model ensemble (Tiny/Base/Large) leverages complementary strengths: Tiny's computational efficiency, Base's balanced representation capacity, and Large's sophisticated feature extraction. This combination reduces variance through diverse receptive field coverage.

The ensemble's success stems from three architectural innovations: 1) DASPPV2's dynamic multi-scale context fusion adapts to varying wound sizes, 2) Transformer context blocks

TABLE I: Comparison of Dice Scores for Different Wound Segmentation Models and Ensembles

Model	Dice Score (%)
Baseline Models	
MobileNetV2 + CCL [14]	90.47
M-Vgg19-U-Net [17]	92.02
LinkNet-EffB1 [16]	92.09
WDAP Linknet [8]	92.70
WoundSegFormerV2 with Different Variants	
Tiny	91.90
Small	91.79
Base	92.43
Large	92.46
Tiny, Small, Large	92.57
Tiny, Small, Base, Large	92.66
Base, Large	**92.76**
Small, Base, Large	**92.78**
Tiny, Base, Large	**92.97**

enhance long-range dependency modeling for irregular boundaries, and 3) CBAM-augmented decoders recover fine spatial details lost during downsampling.

These results establish WoundSegFormerV2 as a new state-of-the-art solution for automated wound segmentation, combining clinical-grade accuracy with operational practicality. The architecture's robustness to dataset noise and lighting variations positions it well for real-world deployment across diverse healthcare settings.

VI. Conclusion

This work presented WoundSegFormerV2, a novel hybrid architecture integrating ConvNeXt-based feature extraction with transformer context modeling for robust wound segmentation. Our approach achieved state-of-the-art performance on the AZH dataset with a Dice score of 92.97% and IoU of 86.95%, significantly outperforming previous methods. The superior results stem from three key innovations: (1) learnable branch weights in DASPPV2 that dynamically adjust multi-scale feature fusion, (2) a global transformer context block that effectively models long-range dependencies crucial for irregular wound boundaries, and (3) CBAM-enhanced decoder pathways that precisely recover spatial details.

Our ensemble strategy effectively leveraged complementary strengths of multiple backbone configurations, with the Tiny/Base/Large combination providing optimal performance. The architecture demonstrates particular robustness to challenging conditions including varying lighting, irregular morphologies, and small lesion detection—factors critical for clinical adoption.

By improving segmentation accuracy and reliability, WoundSegFormerV2 enables more standardized wound documentation and assessment, directly impacting treatment decisions for chronic conditions like diabetic foot ulcers. Future work should focus on extending this framework to multi-class wound segmentation for distinguishing between tissue types and exploring lightweight adaptations for mobile deployment in resource-constrained settings. This work represents a significant advancement toward clinically viable automated wound analysis systems that can standardize care across diverse healthcare environments.

References

[1] C. K. Sen, "Human wound and its burden: Updated 2022 compendium of estimates," *Advances in Wound Care*, vol. 12, no. 12, pp. 657–670, 2023, pMID: 37756368. [Online]. Available: https://doi.org/10.1089/wound.2023.0150

[2] Y. Lei, Y. Fu, T. Wang, R. L. J. Qiu, W. J. Curran, T. Liu, and X. Yang, "Deep learning in multi-organ segmentation," 2020. [Online]. Available: https://arxiv.org/abs/2001.10619

[3] R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, and A. K. Nandi, "Medical image segmentation using deep learning: A survey," *IET Image Processing*, vol. 16, no. 5, p. 1243–1267, Jan. 2022. [Online]. Available: http://dx.doi.org/10.1049/ipr2.12419

[4] Z. Liu, L. Tong, Z. Jiang, L. Chen, F. Zhou, Q. Zhang, X. Zhang, Y. Jin, and H. Zhou, "Deep learning based brain tumor segmentation: A survey," 2021. [Online]. Available: https://arxiv.org/abs/2007.09479

[5] F. Long, "Microscopy cell nuclei segmentation with enhanced U-Net," *BMC Bioinformatics*, vol. 21, no. 1, p. 8, Jan. 2020.

[6] W. Liu, J. Luo, Y. Yang, W. Wang, J. Deng, and L. Yu, "Automatic lung segmentation in chest x-ray images using improved U-Net," *Sci. Rep.*, vol. 12, no. 1, p. 8649, May 2022.

[7] I. Marinakis, K. Karampidis, and G. Papadourakis, "Pulmonary nodule detection, segmentation and classification using deep learning: A comprehensive literature review," *BioMedInformatics*, vol. 4, no. 3, pp. 2043–2106, Sep. 2024.

[8] S. R. Oota, V. Rowtula, S. Mohammed, M. Liu, and M. Gupta, "WSNet: Towards an effective method for wound image segmentation," in *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. IEEE, Jan. 2023.

[9] S. Thomas., "Medetec wound dataset," 2017.

[10] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," 2022. [Online]. Available: https://arxiv.org/abs/2201.03545

[11] P. Gu, Y. Zhang, C. Wang, and D. Z. Chen, "Convformer: Combining cnn and transformer for medical image segmentation," 2022. [Online]. Available: https://arxiv.org/abs/2211.08564

[12] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[13] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," 2018. [Online]. Available: https://arxiv.org/abs/1807.06521

[14] C. Wang, D. M. Anisuzzaman, V. Williamson, M. K. Dhar, B. Rostami, J. Niezgoda, S. Gopalakrishnan, and Z. Yu, "Fully automatic wound segmentation with deep convolutional neural networks," *Sci. Rep.*, vol. 10, no. 1, p. 21897, Dec. 2020.

[15] C. Cui, K. Thurnhofer-Hemsi, R. Soroushmehr, A. Mishra, J. Gryak, E. Dominguez, K. Najarian, and E. Lopez-Rubio, "Diabetic wound segmentation using convolutional neural networks," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, Jul. 2019.

[16] A. Mahbod, G. Schaefer, R. Ecker, and I. Ellinger, "Automatic foot ulcer segmentation using an ensemble of convolutional neural networks," 2022. [Online]. Available: https://arxiv.org/abs/2109.01408

[17] K. O. M. Goh, M. K. Morol, M. J. Hossen, M. A. Al-Jubair, R. I. Rabbi, and N. Fahad, "The next chapter in wound analysis: Introducing a hybrid model for improved segmentation with the help of deep convolutional neural network," *J. Adv. Res. Appl. Sci. Eng. Tech.*, vol. 63, no. 1, pp. 225–239, Mar. 2025.

[18] R. Niri, S. Zahia, A. Stefanelli, K. Sharma, S. Probst, S. Pichon, and G. Chanel, "Wound segmentation with U-Net using a dual attention mechanism and transfer learning," *J Imaging Inform Med*, Jan. 2025.

[19] M. Z. B. Jahangir, S. Akter, M. A. A. Nasim, K. D. Gupta, and R. George, "Deep learning for automated wound classification and segmentation," 2024. [Online]. Available: https://arxiv.org/abs/2408.11064

[20] D. A. Muhtasim, M. I. Pavel, A. Paul, A. W. Shakib, S. S. Gazi, and M. M. Hasan, "Automated wound segmentation using attention mechanism based on enhanced mobilenetv2," in *2024 IEEE 3rd International Conference on Robotics, Automation, Artificial-Intelligence and Internet-of-Things (RAAICON)*, 2024, pp. 96–101.