Transactions on Computational and Scientific Methods | Vo. 5, No. 2, 2025
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

UIM-Net: A New Frontier in Automated Skin Disease Diagnosis

Leandro Vosquez[1], Eloísa Villavicencio[2], Beltrán Abarca[3], Fermín Escudero[4]
[1]University of Cordoba, Cordoba, Spain
[2]University of Cordoba, Cordoba, Spain
[3]University of Cordoba, Cordoba, Spain
[4]University of Cordoba, Cordoba, Spain
*Corresponding author: Leandro Vosquez, Leandro.v0908@gmail.com

Abstract: This paper introduces UIM-Net, a novel network model designed for the segmentation of skin lesions in medical images. Tested on the ISIC 2018 dataset, UIM-Net demonstrates superior performance compared to established models such as UNet, AttU-Net, UNet++, DeepLabV3, and UNeXt. Quantitative analysis reveals that UIM-Net achieves the highest IOU scores and surpasses other models in F1 scores, as illustrated in Figure 3.9. The model's parameter count and computational complexity are significantly lower, making it a lightweight yet efficient choice. Through TOPSIS assessment, UIM-Net earns the highest score, indicating its advanced segmentation capabilities on the ISIC 2018 skin disease dataset. This study underscores the potential of UIM-Net in enhancing dermatological image analysis.

Keywords: Component; Skin Lesion Segmentation, UIM-Net, Medical Image Analysis, Deep Learning.

1.Introduction

In recent years, skin diseases have become one of the fastest-growing diseases in China. There is a wide variety of skin diseases, most of which present irregular shapes in black or brown pathological colors, posing a threat not only to people's life safety but also having a significant impact on appearance. Traditional methods of diagnosing skin diseases mainly rely on the professional knowledge and experience of doctors, making subjective judgments on the shape and color of the lesion area. However, due to the small inter-class differences among skin diseases, diagnosis not only consumes a lot of doctors' energy but also makes it difficult to ensure the accuracy of recognition.

Therefore, artificial intelligence technology represented by deep learning is increasingly applied to the medical diagnosis of skin diseases. Kaymak et al.[1] used multiple classification tasks based on AlexNet to gradually analyze whether skin lesions are melanocytic or non-melanocytic, and whether they are malignant or benign. The experimental evaluation results show that this method can effectively improve classification results. Simonyan[2], based on AlexNet and optimizing the original convolutional network structure, proposed the VGG network model. VGG obtains deeper and more effective features by simply stacking a series of convolutional blocks and pooling layers.

However, due to the imbalance of medical image categories and the limitation of dataset quantities, existing deep learning-based medical image segmentation and classification methods often fail to achieve ideal accurate prediction results due to insufficient training. In addition, the limited equipment resources in actual medical scenarios also make it very important to explore how to make algorithm models more lightweight and low computational complexity. Therefore, this paper conducts research on the segmentation and classification tasks of skin disease medical images based on the problems faced in medical image processing. The key research content of this paper is as follows:

To achieve rapid and accurate automated segmentation of lesion areas in skin disease medical images, this paper proposes a U-shaped skin disease image segmentation network, UIM-Net, based on Inception and Multilayer Perceptron (MLP). The network establishes channels for extracting high and low-frequency feature information in image data through Inception's high and low-frequency information mixing modules, and combines the MLP optimization module to strengthen the communication between the transition information of the network encoder and decoder, enabling the network to refine the edge details of the target area, thereby improving the accuracy of the prediction results. Experimental results show that UIM-Net has certain advantages in the segmentation task of skin disease medical images. On the ISIC 2018 and PH2 datasets, the F1 score of this network is much higher than that of UNet++. In addition, the computational complexity and number of parameters of UIM-Net are lower than most networks, making it a lightweight segmentation network.

2. Related Work

Deep learning has become a crucial tool in medical image analysis, particularly in skin lesion segmentation and classification. Traditional convolutional neural networks (CNNs) such as VGG and ResNet have demonstrated strong feature extraction capabilities but often struggle with high computational complexity and inefficiencies in medical image datasets. He et al. [3] evaluated the performance of VGG19 on complex visual data, showcasing its effectiveness in feature extraction but highlighting its limitations in computational efficiency. Similarly, Hu et al. [4] proposed a multi-scale Transformer architecture for medical image classification, emphasizing the importance of attention mechanisms in capturing fine-grained features. These studies illustrate the necessity of optimizing deep learning models to balance accuracy and efficiency in medical imaging tasks.

Beyond CNN-based approaches, Transformer models have gained attention for their superior performance in medical data analysis. Zhu et al. [5] explored the use of Transformer architectures in NLP-driven privacy solutions for medical records, demonstrating their capability in handling complex dependencies within structured and unstructured data. Wu et al. [6] improved entity extraction using an adaptive attention and feature embedding mechanism based on BERT, which is relevant for enhancing feature representation in medical text and image processing. These advancements in Transformer-based models suggest their potential for improving medical image segmentation by leveraging attention mechanisms and global context understanding.

Medical image analysis also benefits from graph-based approaches, which can model complex relationships between medical data points. Mei et al. [7] introduced collaborative hypergraph networks for disease risk assessment, improving predictive accuracy through multi-source data fusion. Similarly, Gao et al. [8] proposed a multi-channel hypergraph-enhanced model for sequential visit prediction, optimizing patient trajectory forecasting through structured graph representations. These graph-based methodologies provide valuable insights into leveraging relational information in deep learning models, potentially benefiting segmentation tasks in medical imaging.

In the context of object detection for medical image analysis, He et al. [9] investigated the RT-DETR model, highlighting its applicability in detecting key regions of interest with high precision. The integration of such object detection frameworks into segmentation pipelines can enhance lesion localization and boundary refinement. These studies collectively underscore the importance of developing lightweight, high-performance models tailored to medical imaging constraints, aligning with the objectives of UIM-Net in improving skin lesion segmentation through an optimized combination of Inception modules and MLP-based enhancements.

3. Background

Early computer technologies applied to the field of skin lesion image segmentation include edge detection[10], threshold segmentation[11], and region growing, which primarily utilize changes in image grayscale values and pixel value distributions, among other shallow image features, to delineate regions with similar properties within skin images. These traditional skin lesion image segmentation methods can achieve satisfactory results in specific scenarios.

Khare[12] et al. developed a threshold segmentation method based on the median and contrast of wavelet coefficients, which exhibits superior performance in medical images containing various complex types of noise.

Rahil et al.[13] optimized edge detection effects by determining the optimal image channels, a method that has certain advantages in terms of accuracy, sensitivity, and boundary error.

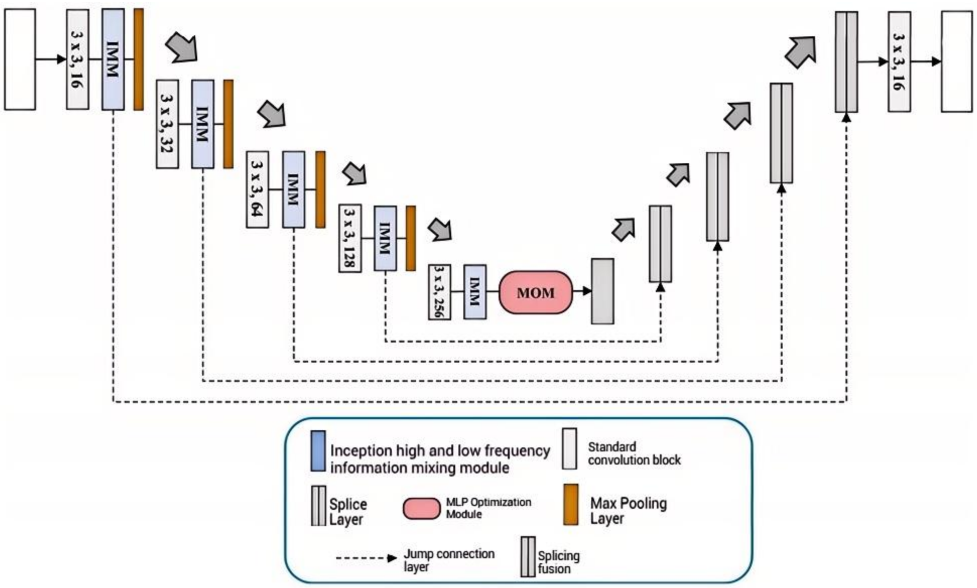

Figure 1. Architecture diagram of UIM-Net

Noorul et al.[14] optimized the seed selection method based on the Harris corner detection theory, proposing an improved region growing segmentation algorithm that offers reliable and robust performance in extracting boundaries and objects.

Glaister et al.[15] proposed a segmentation algorithm based on skin texture, which distinguishes between pathological areas and normal skin by capturing metric differences between textures.

Dang et al.[16] proposed an adaptive threshold method based on color models, demonstrating advanced performance in segmentation tests on the ISIC dataset.

Peruch et al.[17] proposed a new segmentation method for skin lesion areas, MEDS (Mimicking Expert Dermatologists' Segmentations). This method begins with principal component analysis of the color histogram, followed by image preprocessing to reduce noise, and finally, a novel thresholding algorithm clusters pixels in the original image to achieve precise segmentation of skin lesion images. Although these traditional image segmentation algorithms can quickly label images without the need for extensive data learning, they only utilize primary image information, leading to poor performance in skin lesion area segmentation, such as poor continuity, under-segmentation, and low robustness.

In recent years, deep learning algorithms have gradually become the mainstream method in medical image segmentation. The advantage of deep learning algorithms lies in their ability to efficiently and autonomously learn abstract information and intrinsic patterns from data, compared to traditional methods, without the need for complex manual parameter tuning, offering stronger accuracy and robustness. Jonathan et al.[18] constructed an end-to-end operating fully convolutional network (FCN), applying convolutional neural networks heuristically at the pixel level for segmentation, classifying each pixel in the image to achieve the purpose of image segmentation.

4. Method

4.1 Overall Model Structure

UIM-Net's design philosophy is based on the encoder-decoder architecture of UNet. By incorporating the proposed Inception High-Low Frequency Information Mixing module, it significantly compensates for the limitations of convolutions in capturing global information and effectively synthesizes both local and global information from the data. Furthermore, UIM-Net feeds the output features from the deepest layer of the encoder into a key MLP optimization module, further enhancing the capture of spatial information in the network's deep features and strengthening the network's data fitting capability.

Figure 1 illustrates the detailed structure of the UIM-Net model. The input image is processed through both encoder and decoder operations before producing the final output. The encoder stage comprises 5 downsampling processes, where each downsampling halves the feature dimensions, while the decoder stage doubles the feature resolution with each upsampling operation. Notably, the convolutional blocks designed in this paper contain only one convolutional layer in addition to batch normalization operations and ReLU activation units. This design choice aims to balance segmentation accuracy with computational cost, achieving an optimal structure for a lightweight segmentation network. Consequently, UIM-Net has significantly fewer model parameters compared to other segmentation networks like UNet. The channel numbers for each convolutional block are hyperparameters, set from shallow to deep layers as 16, 32, 64, 128, and 256 in this paper.

In the encoder stage, each downsampling process includes standard convolution, Inception High-Low Frequency Information Mixing, and max pooling layers. Image data first undergoes preliminary feature extraction through the convolutional block, then enters the Inception High-Low Frequency Information Mixing module for further refinement of local and global information, and finally passes through a max pooling layer with a stride of 2 to reduce spatial dimensions and select salient information.

The decoder structure mirrors the encoder. Upsampling operations gradually restore the original resolution of the underlying feature maps using bilinear interpolation algorithms and incorporate skip connections to merge feature information of the same level from the encoder, compensating for information lost during the downsampling process. After 5 rounds of information fusion and upsampling, the feature maps are restored to the same dimensions as the original image. In the final part of the decoder, the feature maps undergo point convolution to reduce the number of channels, producing the final prediction results.

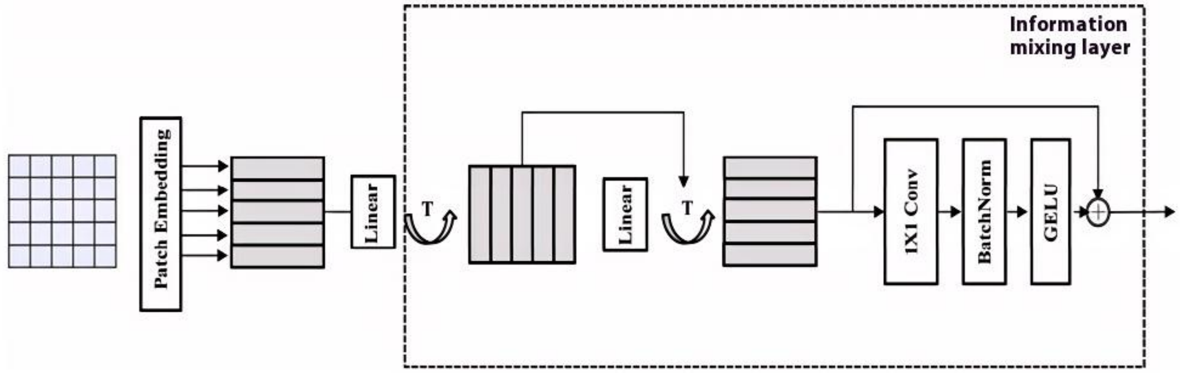

Figure 2. Architecture diagram of MLP optimization module

4.2 MLP Optimization Module

The encoder's primary function is to progressively reduce feature spatial dimensions, increase receptive fields, and extract features. The decoder achieves image segmentation by gradually restoring image size and performing pixel classification. The transition module between the encoder and decoder serves as a crucial structure guiding the transition of high-resolution features from encoder to decoder, using target information obtained from the encoder as guidance to input into the decoder for distinguishing features related to target regions. This significantly impacts the network's ability to correctly locate lesion areas. Therefore, enhancing feature information during this process can improve the network's final segmentation accuracy. This section employs an MLP-based optimization module (MOM) to strengthen information communication within the transition module. This module supports simultaneous interaction between two input dimensional data, enhancing the network's ability to acquire feature map information. The detailed structure of MOM is shown in Figure 2.

Tthe specific steps of MOM can be summarized in three points:

1) First, the input feature map $X_{\text{mequ}} \in R^{(C\times H\times W)}$ is divided into N non-overlapping patches of fixed size $X_p \in RN \times (P^2 \cdot C)$.

2) To further enhance feature extraction capability, each feature patch is serialized into S and mapped to a fixed D dimension through hidden layers, i.e., $R^s \rightarrow R^d$.

3) A standard fully connected layer and a pointwise convolution layer are set up respectively for flexible implementation of channel domain and spatial domain information interaction with the serialized features.

Both the fully connected layer and pointwise convolution operations are followed by GELU activation functions for non-linear activation, enhancing the model's expressiveness.

MOM facilitates better effective information acquisition during feature extraction by employing a fully connected layer and pointwise convolution operation for channel domain and spatial domain information interaction respectively. Compared to the MLP Mixer proposed by Tolstikhin[46], the MOM proposed in this paper maintains a similar architecture but includes three key optimizations:

1) Uses only one fully convolutional layer for channel domain information interaction, reducing parameters and computational complexity while maintaining accuracy.

2) Employs pointwise convolution instead of multilayer perceptrons for spatial domain information interaction.

3) Uses smaller patch sizes (4×4) when dividing patches, significantly reducing the model's computational load.

4.3 Inception High-Low Frequency Information Mixing Module

The IMM module consists of three parts: maximum pooling branch, depthwise separable convolution branch, and dilated convolution branch. Similar to GoogleNet[61], IMM utilizes max pooling and depthwise separable convolution to extract local information, while using dilated convolution to expand the receptive field for global information modeling. This enhances the comprehensiveness of information extraction during the convolution process and reduces the possibility of feature loss. The detailed structure of the IMM module is shown in Figure 3.

1) Maximum Pooling Operation Branch

The maximum pooling operation branch serves to select more distinctive features while preserving more local information. First, the max pooling branch uses a 3×3 pooling kernel with stride 1 to extract maximum values from the input feature map's receptive field, followed by pointwise convolution mapping to obtain Y_{h1}. The max pooling operation can be expressed as Formula 1:

$$Y_{h1} = FC(\mathrm{MaxPool}\ (X_{h1}))$$

where $f(x, y)$ represents the input convolution features, MaxPool refers to the maximum pooling operation, FC refers to pointwise convolution for further depth-wise expansion and combination. Y_{k1} is the value after this branch's operation, used for fusion with outputs from the other two branches in the final stage of the Inception high-low frequency information mixing module.

2) Depthwise Separable Convolution Branch

Depthwise separable convolution[58] can be decomposed into depthwise convolution and pointwise convolution. Depthwise convolution is a channel-based convolution method that applies different convolution kernels to feature maps of each channel, reconstructing them into new multi-channel feature maps. However, since channel connections are blocked, spatial information cannot be obtained, thus pointwise convolution is used to recombine spatial information. Through these two decomposition processes, depthwise separable convolution can achieve detail perception with fewer computational resources. Notably, this section uses two fully connected layers and a GELU activation unit to form pointwise convolution to enhance the network's sensitivity to feature information. The mathematical expression for the depthwise separable convolution branch is shown in Formula 2:

$$Y_{h2} = \mathrm{Dw\ Conv}\ \ (FC(X_{h2}))$$

3) Dilated Convolution Branch

Standard convolution ignores the construction of long-distance information dependencies and only focuses on local information, which is why networks like UNet cannot improve accuracy further. Increasing the receptive field, i.e., expanding the spatial range of element mapping on feature maps, can obtain multi-scale contextual information and compensate for the lack of global information in convolution. Based on this background, Chen et al.[62] proposed DeepLabv3, which achieved significant improvement in segmentation performance through dilated convolution and achieved excellent results in segmentation tasks. As the name suggests, dilated convolution fills gaps in standard convolution kernels to increase the convolution receptive field range. Inspired by this, this branch introduces dilated convolution for weighted computation of input feature elements to enable global information acquisition capability. Notably, different dilation rates are applied in different IMM dilated convolution branches in this paper, specifically 2, 4, 8, 16, and 32.

The computation process of IMM can be summarized in Formula 3:

$$Y_{prm} = f_{3\times3}\{[f_{1\times1}(F_{mp}(X)), F_{dc}(f_{1\times1}(X))]\} + X \quad (3)$$

where $f_{3\times 3}$ and $f_{1\times 1}$ are standard 3×3 and 1×1 convolutions respectively, including convolution, batch normalization[63], and ReLU activation unit; + represents element-wise addition.

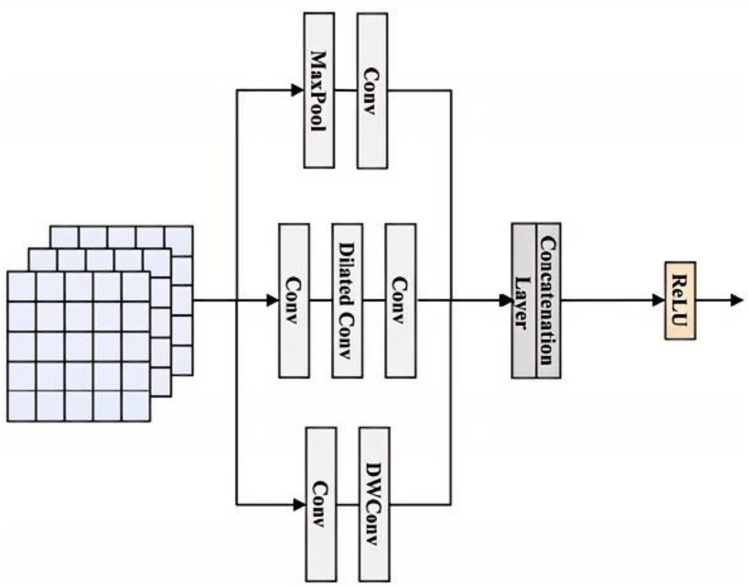

Figure 3. Architecture diagram of Inception high and low frequency information hybrid module

5. Experiment

5.1 Dataset

The dermatological medical image segmentation dataset used in this paper is the ISIC 2018 dataset[60], which is a publicly available dataset released by the International Skin Imaging Collaboration (ISIC). This dataset includes 2,594 training images and 100 test images of skin disease patients, along with their corresponding label images.

5.2 Baselines

In this paper, six classic network models used for medical image segmentation are selected and compared with the proposed UIM-Net from multiple dimensions, including accuracy and prediction efficiency. The comparative segmentation network models are as follows: UNet1111, UNet++2222, DeepLabV36262, UNeXt2929, ResUNet6666.

5.3 Experiment Results Analysis

This section tests the UIM-Net network model proposed in this paper on the ISIC 2018 dataset. Furthermore, to demonstrate the performance advantages of UIM-Net, it is compared with the most classic models in the field of medical image segmentation (UNet, AttU-Net, UNet++, DeepLabV3, UNeXt). The following analysis of UIM-Net is conducted from both quantitative and qualitative perspectives.

Table 1. Experiment Results

Method	F1	IOU	G-mean	TOPSIS
Unet	87.21	79.15	92.54	17.70
AttU-Net	85.69	76.95	92.27	11.20
UNet++	83.04	73.95	90.08	0

Method	F1	IOU	G-mean	TOPSIS
DeepLabV3	88.15	80.50	93.08	21.78
UNeXt	88.71	81.22	93.7	24.10
UIM-Net	88.85	81.14	94.29	25.21

The experimental quantitative results are shown in Table 1, where the UIM-Net proposed in this paper achieved good performance across several metrics. Notably, compared to UNet and UNet++, UIM-Net obtained the highest IOU scores. Tthe F1 scores also surpassed other models, with a significant effect. It is worth noting that UIM-Net's number of parameters and computational complexity also hold certain advantages among all network models. Additionally, this subsection employs the TOPSIS method to comprehensively evaluate the aforementioned F1, IOU, and G-mean scores, obtaining precise scores that reflect the differences between various models. It is evident that UIM-Net has the highest TOPSIS score. This indicates that UIM-Net possesses advanced segmentation capabilities on the ISIC 2018 skin disease dataset.

6. Conclusion

The UIM-Net model has proven to be a significant advancement in the field of dermatological image segmentation. Through rigorous testing on the ISIC 2018 dataset, it has outperformed several benchmark models in terms of accuracy and efficiency. With its superior IOU and F1 scores, UIM-Net not only demonstrates its capability to achieve high segmentation accuracy but also its advantage in terms of parameter economy and computational efficiency. The TOPSIS assessment further solidifies UIM-Net's standing as a top-performing model, highlighting its potential for practical application in clinical settings. This study concludes that UIM-Net is a promising tool for enhancing the accuracy and efficiency of skin lesion segmentation, contributing to the advancement of dermatological diagnostics.

References

[1] Kaymak S, Esmaili P, and Serener A, "Deep learning for two-step classification of malignant pigmented skin lesions," Proceedings of the 2018 14th Symposium on Neural Networks and Applications (NEUREL), pp. 1-6, 2018.

[2] Simonyan K and Zisserman A, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[3] W. He, T. Zhou, Y. Xiang, Y. Lin, J. Hu, and R. Bao, "Deep learning in image classification: Evaluating VGG19’s performance on complex visual data," arXiv preprint arXiv:2412.20345, 2024.

[4] J. Hu, Y. Xiang, Y. Lin, J. Du, H. Zhang, and H. Liu, "Multi-Scale Transformer Architecture for Accurate Medical Image Classification," arXiv preprint arXiv:2502.06243, 2025.

[5] Z. Zhu, Y. Zhang, J. Yuan, W. Yang, L. Wu, and Z. Chen, "NLP-Driven Privacy Solutions for Medical Records Using Transformer Architecture," unpublished.

[6] L. Wu, J. Gao, X. Liao, H. Zheng, J. Hu, and R. Bao, "Adaptive Attention and Feature Embedding for Enhanced Entity Extraction Using an Improved BERT Model," unpublished.

[7] T. Mei, Z. Zheng, Z. Gao, Q. Wang, X. Cheng, and W. Yang, "Collaborative Hypergraph Networks for Enhanced Disease Risk Assessment," Proceedings of the International Conference on Electronics, Data, and Computational Science (ICEDCS), pp. 416-420, Sep. 2024.

[8] Z. Gao, T. Mei, Z. Zheng, X. Cheng, Q. Wang, and W. Yang, "Multi-Channel Hypergraph-Enhanced Sequential Visit Prediction," Proceedings of the International Conference on Electronics, Data, and Computational Science (ICEDCS), pp. 421-425, Sep. 2024.

[9] W. He, Y. Zhang, T. Xu, T. An, Y. Liang, and B. Zhang, "Object detection for medical image analysis: Insights from the RT-DETR model," arXiv preprint arXiv:2501.16469, 2025.

[10] Canny J, "A computational approach to edge detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, no. 6, pp. 679-698, 1986.

[11] Reddi S. S, Rudin S. F, and Keshavan H. R, "An optimal multiple threshold scheme for image segmentation," IEEE Transactions on Systems, Man, and Cybernetics, no. 4, pp. 661-665, 1984.

[12] Khare A and Tiwary U. S, "Soft-thresholding for denoising of medical images—a multiresolution approach," International Journal of Wavelets, Multiresolution and Information Processing, vol. 3, no. 4, pp. 477-496, 2005.

[13] Garnavi R, Aldeen M, Celebi M. E, et al., "Border detection in dermoscopy images using hybrid thresholding on optimized color channels," Computerized Medical Imaging and Graphics, vol. 35, no. 2, pp. 105-115, 2011.

[14] X. Yan, W. Wang, M. Xiao, Y. Li, and M. Gao, "Survival prediction across diverse cancer types using neural networks", Proceedings of the 2024 7th International Conference on Machine Vision and Applications, pp. 134-138, 2024.

[15] Glaister J, Wong A, and Clausi D. A, "Segmentation of skin lesions from digital images using joint statistical texture distinctiveness," IEEE Transactions on Biomedical Engineering, vol. 61, no. 4, pp. 1220-1230, 2014.

[16] M. Xiao, Y. Li, X. Yan, M. Gao, and W. Wang, "Convolutional neural network classification of cancer cytopathology images: taking breast cancer as an example," Proceedings of the 2024 7th International Conference on Machine Vision and Applications, pp. 145–149, Singapore, Singapore,2024

[17] W. Wang, Y. Li, X. Yan, M. Xiao and M. Gao, "Breast cancer image classification method based on deep transfer learning," Proceedings of the International Conference on Image Processing, Machine Learning and Pattern Recognition, pp. 190-197, 2024.

[18] Long J, Shelhamer E, and Darrell T, "Fully convolutional networks for semantic segmentation," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431-3440, 2015.