melSeg: An Adaptation of Segment Anything Model for Skin Lesion Segmentation

1st Shudipto Sekhar Roy
Department of Mechanical Engineering
University of North Dakota
Grand Forks, USA
shudipto.sekharroy@und.edu

2nd Ramtin Kardan
Department of Mechanical Engineering
University of North Dakota
Grand Forks, USA
ramtin.kardan@und.edu

3rd Jeremiah Neubert
Department of Mechanical Engineering
University of North Dakota
Grand Forks, USA
jeremiah.neubert@und.edu

2025 IEEE 15th Annual Computing and Communication Workshop and Conference (CCWC) | 979-8-3315-0769-5/25/$31.00 ©2025 IEEE | DOI: 10.1109/CCWC62904.2025.10903695

***Abstract*—Skin cancer, particularly melanoma, is a major global health concern. Numerous digital analysis methods are available for detecting malignant skin lesions; however, limited accuracy in lesion identification can reduce their effectiveness in providing reliable diagnoses. This study presents a deep learning approach to develop an image segmentation model, melSeg, which accurately segments skin lesion areas and enhances diagnostic accuracy. We employ the state-of-the-art Segment Anything Model (SAM), originally designed for general-purpose segmentation tasks. To adapt SAM for skin lesion segmentation, we incorporate a feature pyramid network (FPN) to capture lesion features at multiple scales during fine-tuning. The model was rigorously evaluated on the HAM10000 dataset, a publicly available resource containing 10,015 dermatoscopic images of skin lesions annotated with corresponding lesion area masks. On this dataset, our adapted SAM model achieved an Intersection over Union (IoU) score of 89.29% and a Dice score of 94.19%, marking a substantial improvement over contemporary state-of-the-art methods in skin lesion segmentation. These results underscore the potential of our model in early skin cancer detection, offering more accurate diagnostic tools to aid in combating skin cancer.**

Keywords: deep learning, segment anything model, image segmentation, skin lesion, HAM10000

I. Introduction

Skin cancer, especially malignant melanoma, is a growing global health crisis with increasing incidence and associated mortality rates. In 2018 alone, approximately 1.3 million new cases of skin cancer were reported globally, including over 287,000 cases of melanoma, leading to nearly 61,000 deaths [1]. Statistics also show that in the United States, over 207,000 cases of skin cancer were reported in 2021, with melanoma accounting for a significant portion [2]. Projections indicate that newly diagnosed melanoma cases will increase by over 50% by 2040, underscoring the urgent need for improved diagnostic and treatment strategies [3]. The stage at which melanoma is detected greatly influences patient outcomes; early-stage diagnosis often results in favorable prognoses, whereas advanced metastatic melanoma is associated with a dismal 5-year survival rate of less than 10% [4]. This alarming trend highlights the critical importance of developing early intervention strategies to mitigate the rising burden of melanoma-related deaths.

Dermoscopy is a foundational tool in the early detection of melanoma. This non-invasive imaging technique enhances the visualization of pigmented skin lesions, allowing dermatologists to detect melanomas that might not be visible to the naked eye. By enhancing features like pigment patterns and vascular structures, dermoscopy aids in identifying melanoma in its early stages. Traditionally, dermatologists have relied on the ABCDE rule to assess skin lesions, evaluating characteristics like asymmetry, border irregularity, color variation, diameter, and lesion evolution as potential indicators of melanoma [5]. Despite these advances, manual evaluation of dermoscopic images remains time-consuming and prone to inconsistency among clinicians. In contrast, computer-aided diagnostic systems offer a promising alternative, providing enhanced accuracy and efficiency. However, precise segmentation of skin lesions remains challenging due to variability in lesion size, shape, color, and the presence of artifacts such as hair.

Machine learning is increasingly applied to detect objects of interest in images or predict disease or event probabilities using historical data. Leveraging advanced algorithms and large datasets, machine learning improves decision-making processes and outcomes across multiple domains [6-8]. Applying machine learning techniques to dermatologic imaging has the potential to transform skin cancer diagnostic procedures. Research suggests that machine learning algorithms can reduce the number of artifacts needing review, expedite the diagnostic process, and decrease patient clinic visits [9]. The literature on skin cancer detection reflects the evolution of methodologies, from basic image processing to advanced machine learning-based approaches, yielding substantial improvements in both classification and segmentation tasks.

Deep learning, in particular, has driven major advancements in medical image analysis, leading to significant gains in the segmentation and classification of skin lesions. Convolutional Neural Networks (CNNs) have emerged as powerful tools for extracting hierarchical features from complex medical images, with numerous studies leveraging these networks to push the state-of-the-art in melanoma detection. For example, Afza et al. [10] developed a hierarchical framework involving preprocessing, feature extraction, and classification phases. They used a fine-tuned ResNet50 model and an optimization algorithm, achieving improved classification accuracy across multiple dermoscopy datasets. Similarly, Al-Masni et al. [11] introduced a two-stage deep learning framework, utilizing a fully resolved CNN (FrCNN) for lesion segmentation, followed by classification with multiple pre-trained networks, demonstrating the effectiveness of

integrating different deep learning architectures for enhanced outcomes.

Building on this foundation, our current work focuses on improving skin lesion segmentation by adapting general-purpose deep learning models for specialized tasks. In our previous research, we employed models such as YOLOv2 for real-time melanoma detection [12] and the Inception-v3 deep learning model to classify melanoma in both dermoscopic and digital images [13]. In this study, we present an enhanced segmentation model trained on the HAM10000 dataset, a widely used benchmark in skin lesion analysis [14]. Our approach utilizes the Segment Anything Model (SAM) [15] to train a robust segmentation model. SAM is designed to perform semantic segmentation by effectively learning to identify and differentiate between distinct regions of interest. By leveraging this capability, our model accurately segments specific lesion areas, providing a precise delineation crucial for subsequent analysis. To enhance feature extraction, we employed the Feature Pyramid Networks (FPN) technique [16]. FPNs improve deep learning models' ability to detect objects across various scales by constructing a multi-level feature pyramid from a single-scale input. FPNs enhance accuracy in object detection and segmentation tasks by producing high-resolution feature maps enriched with robust semantic information through a top-down pathway with lateral connections. To contextualize our approach, we compare our results with recent advancements in skin lesion analysis, highlighting contributions from other works in the field, as summarized below.

In this section, we focus on studies utilizing the HAM10000 dataset to facilitate a comparative analysis of the results. Chu et al. [17] introduced a multi-task learning approach that jointly trains a segmentation and classification model using RECIST measurements. MFSNet is another novel deep-learning framework for supervised skin lesion segmentation [18]. This model integrates multi-scale feature maps to generate accurate segmentation masks using a Res2Net backbone combined with attention modules. Yang et al. [19] introduced Rema-Net, a lightweight multi-attention CNN designed for efficient skin lesion segmentation, which improved segmentation performance while using almost 40% fewer parameters than U-Net by combining spatial and reverse attention mechanisms. A hybrid method for skin lesion analysis that combines deep learning techniques with conventional image processing was proposed by Bibi et al. [20]. The method involves contrast enhancement using a fusion of filtering techniques, lesion segmentation, and feature extraction using MobileNet V2 and VGG16 pre-trained models. A cubic SVM was used for classification after features were refined using a maximum entropy score-based selection (MESbS) and fused using canonical correlation analysis. Karri et al. [21] presented a two-phase cross-domain transfer learning approach to improve skin lesion segmentation. Their technique introduced a deep learning model that enhances segmentation accuracy by using spatial edge attention fusion and receptive field blocks. This study also provided quantitative results for skin lesion segmentation on the HAM10000 dataset using multiple other conventional methods.

II. Methodology

This research can be divided into four major steps. First, the HAM10000 dataset, a comprehensive skin lesion dataset, was sourced from a publicly available domain. Next, the Segment Anything Model (SAM) was adapted to enhance segmentation performance by integrating a Feature Pyramid Network (FPN), enabling better feature extraction at multiple scales. The adapted model was then trained on the HAM10000 dataset using a five-fold cross-validation technique to ensure robust evaluation. Finally, during the evaluation phase, multiple metrics, including Dice Score and Intersection over Union (IoU), were employed to assess the model's segmentation performance. Fig. 1 illustrates the research methodology in a flow diagram.

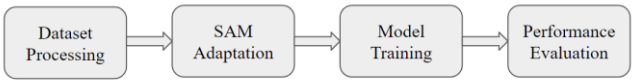

Fig. 1. Methodology of the proposed work.

A. Dataset Analysis

The HAM10000 dataset consists of 10,015 high-quality dermoscopic images, captured in a clinical setting with zoomed-in details. These images were gathered over nearly 20 years and include samples from individuals of different ages, covering various locations on the body. This dataset includes skin lesions from seven diagnostic categories, with 6,705 images representing melanocytic nevi, resulting in a skew toward this category. Since this research focuses on skin lesion segmentation, addressing this imbalance was considered outside the scope. Each image has a fixed size of 450x600 pixels and is paired with a corresponding skin lesion segmentation mask of the same dimensions. Fig. 2 presents three sample images from this dataset along with their corresponding segmentation masks.

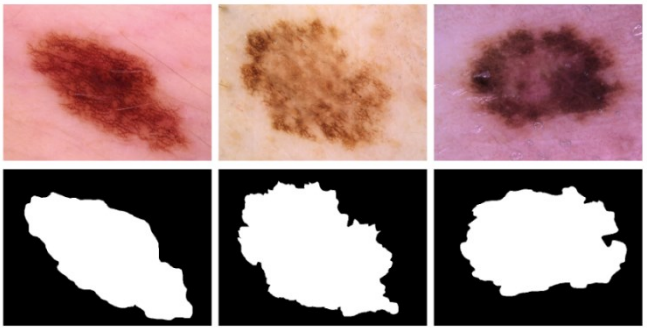

Fig. 2. Samples from the HAM10000 dataset with corresponding masks.

B. Adaptation of Segment Anything Model

In this research, we developed a deep learning model for skin lesion segmentation based on the Segment Anything Model (SAM). SAM consists of three main components: an image encoder that encodes the input image, a prompt encoder that encodes given prompts, and a mask decoder that takes the embedded outputs of the image encoder and prompt encoder to produce segmentation masks. The image encoder leverages the architecture of a Vision Transformer (ViT) to establish a self-attention mechanism across grids in the input image, capturing both local and global context effectively. Fig. 3 illustrates the basic working principle of the Segment Anything Model.

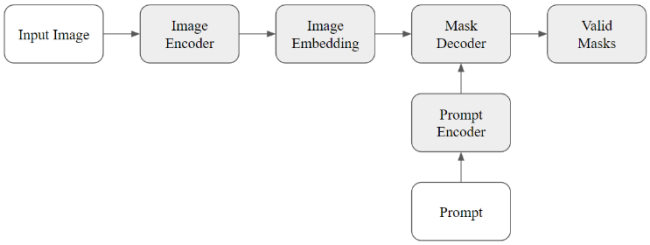

Fig. 3. Diagram of Segment Anything Model (SAM).

Although SAM is robust in capturing the spatial context of an image, the dataset we are working with consists of close-up, high-resolution images focused tightly on the skin-lesion area. This setup provides limited context beyond the lesion itself, as the primary objective is to highlight lesion details rather than the broader surroundings. To overcome this limitation and fully leverage SAM's robustness, we integrated a Feature Pyramid Network (FPN) just before SAM's image encoding stage. Fig. 4 demonstrates this adaptation of SAM for our study.

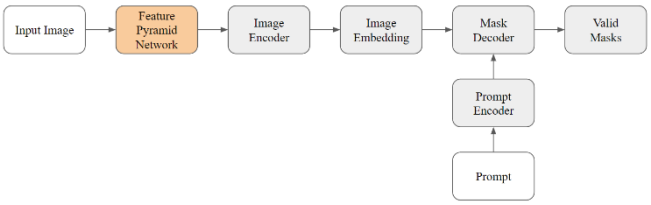

Fig. 4. Diagram of the proposed adaptation of Segment Anything Model.

This FPN enhances feature extraction from the input image by capturing information at multiple spatial scales. Fig. 5 illustrates the two main components of an FPN: the bottom-up and top-down sections. In the bottom-up section, the input image is processed from lower to higher levels of the network, generating progressively higher levels of semantic information at each layer. In the top-down section, the feature maps are upsampled and combined with their corresponding maps from the bottom-up section.

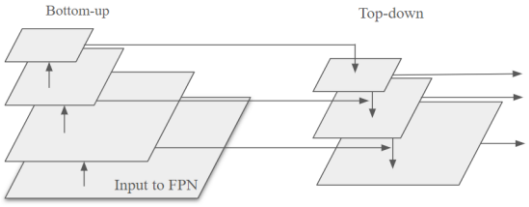

Fig. 5. Diagram of a Feature Pyramid Network (FPN).

C. Model Training

The proposed model was trained using five-fold cross-validation, where the dataset was divided into five equal parts. In each iteration, the model was trained on four folds and validated on the remaining one. This approach helps ensure that the model's performance is evaluated on diverse subsets of data, reducing the risk of overfitting. The final performance metric was calculated as the average of results across all five folds

This research explored two loss functions: DiceCELoss and FocalLoss. DiceCELoss combines Dice Loss and Cross-Entropy Loss (CE Loss) to mitigate each loss's individual limitations, while FocalLoss addresses class imbalance by reducing the impact of well-classified pixels and emphasizing harder-to-classify, misclassified pixels, helping the model focus on challenging areas.

The optimizers Adam and SGD were tested with learning rates of 1e-4, 1e-5, and 1e-6. This combination of variations resulted in twelve experiments, all conducted on an Nvidia GeForce RTX 4090 with 24GB of memory. Due to hardware limitations, a fixed batch size of four was maintained across all experiments.

D. Performance Evaluation

The model's performance was evaluated using the following five metrics: precision (1), recall (2), accuracy (3), Dice Score (4), and IoU (5). Each metric is derived from the counts of True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN). Training was conducted using a five-fold cross-validation approach, and the average of these metrics on the validation set was used to assess the model's performance.

$$Precision = TP/(TP + FP) \quad (1)$$

$$Recall = TP/(TP + FN) \quad (2)$$

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (3)$$

$$Dice\ Score = 2TP/(2TP + FP + FN) \quad (4)$$

$$IoU = TP/(TP + FP + FN) \quad (5)$$

As precision, recall, and accuracy metrics were found to be potentially misleading for this segmentation task, we prioritized optimizing performance based on Dice Score and IoU metrics. Fig. 6 demostrates five examples from the model's inference pipeline, where samples from the validation dataset were processed by the trained model. The output mask (highlighted in yellow) and ground-truth mask (highlighted in green) are overlaid on the input image to visualize the model's performance.

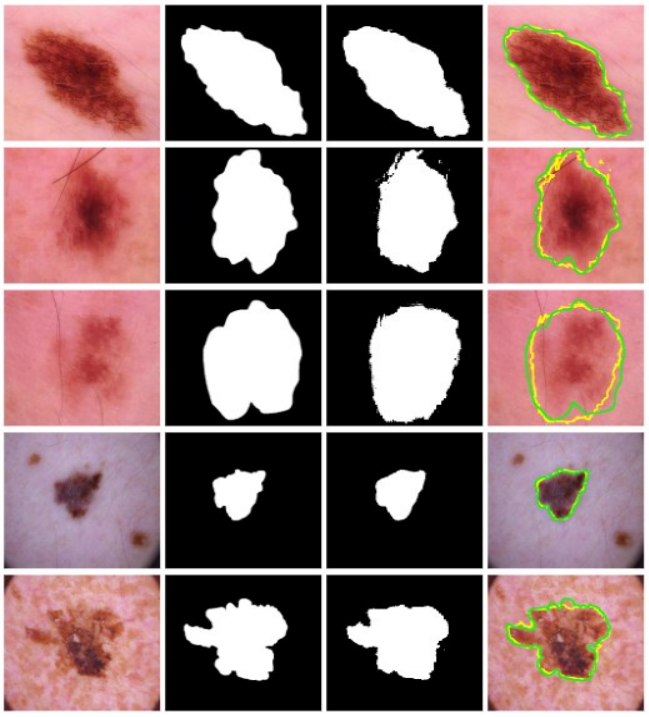

Fig. 6. Column (a) displays the input images; column (b) presents the ground-truth masks; column (c) shows the prediction masks; and column (d) overlays the ground-truth in green and the predictions in yellow on the input images.

III. Result and Discussion

This section presents the performance evaluation of the proposed model. A total of twelve experiments were conducted, with various hyperparameters adjusted to optimize results. Due to the computationally intensive architecture of the Segment Anything Model, the hardware used permitted a maximum batch size of four during training. Multiple metrics were calculated to assess model performance; however, for the image segmentation task, Intersection over Union (IoU) and Dice Score were selected as the primary metrics to identify the best-performing experiment. Each of the twelve experiments underwent a five-fold cross-validation process to produce the reported metrics, ensuring robust and reliable results. Table 1 displays the results from all twelve experiments, with the first row demonstrating the highest performance, achieving a Dice Score of 94.19% and an IoU of 89.29%.

Table I. Average Results from Five-fold Cross Validation with Batch Size of 4

Obs.	Configuration	Accu-racy	Preci-sion	Recall	Dice-Score	IoU
01	L.R.: 1e-4 Loss Func.: DiceCELoss Optim.: Adam	96.62 %	94.32 %	94.30 %	94.19 %	89.29 %
02	L.R.: 1e-5 Loss Func.: DiceCELoss Optim.: Adam	96.62 %	94.42 %	94.12 %	94.17 %	89.24 %
03	L.R.: 1e-6 Loss Func.: DiceCELoss Optim.: Adam	96.38 %	93.88 %	93.83 %	93.74 %	88.52 %
04	L.R.: 1e-4 Loss Func.: FocalLoss Optim.: Adam	96.59 %	94.23 %	94.22 %	94.11 %	89.17 %
05	L.R.: 1e-5 Loss Func.: FocalLoss Optim.: Adam	96.49 %	94.30 %	93.79 %	93.93 %	88.82 %
06	L.R.: 1e-6 Loss Func.: FocalLoss Optim.: Adam	96.26 %	93.31 %	93.93 %	93.50 %	88.11 %
07	L.R.: 1e-4 Loss Func.: DiceCELoss Optim.: SGD	96.32 %	93.45 %	94.06 %	93.63 %	88.28 %
08	L.R.: 1e-5 Loss Func.: DiceCELoss Optim.: SGD	95.90 %	92.67 %	93.50 %	92.94 %	87.09 %
09	L.R.: 1e-6 Loss Func.: DiceCELoss Optim.: SGD	95.05 %	91.12 %	92.46 %	91.59 %	84.78 %
10	L.R.: 1e-4 Loss Func.: FocalLoss Optim.: SGD	95.89 %	92.72 %	93.32 %	92.88 %	86.98 %
11	L.R.: 1e-5 Loss Func.: FocalLoss Optim.: SGD	95.34 %	91.82 %	92.41 %	91.95 %	85.39 %
12	L.R.: 1e-6 Loss Func.: FocalLoss Optim.: SGD	93.67 %	89.39 %	89.10 %	88.96 %	80.48 %

Table II. Comparative Analysis: MelSeg Model Versus State-of-the-Art Models on HAM10000 Dataset

Model	IoU	Dice Score
A1+L [17]	83.51%	90.54%
MFSNet [18]	90.20%	90.60%
Rema-Net [19]	88.81%	93.59%
AttentionUnet [21]	-	88.42%
LEDNet [21]	-	84.65%
melSeg	**89.29%**	**94.19%**

Table 2 compares the performance of the proposed model with contemporary state-of-the-art methods for skin lesion segmentation on the HAM10000 dataset. For the studies present by [21], IoU values were not reported and are therefore left blank. The results show that our proposed model outperforms others in Dice Score while maintaining an impressive IoU. Although MFSNet [18] marginally outperforms the model in IoU, it is significantly surpassed in Dice Score, demonstrating the proposed model's superior ability to accurately segment skin lesions.

This advantage in Dice Score, when compared to other contemporary methods, highlights the model's ability to achieve more precise delineation of lesion boundaries, which is crucial for accurate diagnostic analysis. Overall, the model's performance underscores its potential as an effective tool for advancing early skin cancer detection.

IV. Conclusion and Future Work

This research aimed to develop a deep learning model capable of accurately segmenting skin lesion areas from digital images, with the goal of providing a more effective tool for skin cancer detection. The proposed model is an adaptation of the renowned Segment Anything Model (SAM), enhanced by the incorporation of a feature pyramid network (FPN) to capture lesion features at multiple scales during fine-tuning. This multi-scale approach allows the model to effectively manage variations in lesion size and structure, which are often challenging in medical image analysis. Training was conducted on the HAM10000 dataset using a fixed batch size of four due to hardware constraints. To thoroughly assess the model's performance, a five-fold cross-validation approach was employed, and the results indicate the model's effectiveness in meeting the research objectives. The model achieved an Intersection over Union (IoU) of 89.29% and a Dice Score of 94.19%, outperforming several state-of-the-art methods on the HAM10000 dataset. These findings highlight the potential of this model to improve skin cancer detection classifiers by accurately segmenting lesion areas, enabling focused diagnostic analysis on regions of interest.

In future work, we plan to assess the model's performance on additional skin lesion datasets, such as PH2 [22] and MEDNODE [23]. These datasets will provide a broader range of lesion types, demographic variations, and differing image qualities. Evaluating the model on these datasets will offer valuable insights into its ability to handle diverse real-world scenarios and variations in input data.

REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA: a cancer journal for clinicians,* vol. 69, no. 1, pp. 7-34, 2019.

[2] R. Siegel, K. Miller, and R. A. Jemal, "Cancer facts & figures 2021," 2021.

[3] M. Arnold *et al.*, "Global burden of cutaneous melanoma in 2020 and projections to 2040," *JAMA dermatology,* vol. 158, no. 5, pp. 495-503, 2022.

[4] L. Bomar, A. Senithilnathan, and C. Ahn, "Systemic therapies for advanced melanoma," *Dermatologic clinics,* vol. 37, no. 4, pp. 409-423, 2019.

[5] G. M. Shahriar Himel, M. Islam, K. Abdullah Al-Aff, S. Ibne Karim, and M. K. U. Sikder, "Skin Cancer Segmentation and Classification Using Vision Transformer for Automatic Analysis in Dermatoscopy-based Non-invasive Digital System," *arXiv e-prints,* p. arXiv: 2401.04746, 2024.

[6] R. Kardan, S. S. Roy, A. Elsaharti, and J. Neubert, "Machine Learning Based Specularity Detection Techniques To Enhance Indoor Navigation," in *2023 IEEE 17th International Conference on Semantic Computing (ICSC)*, 2023: IEEE, pp. 143-148.

[7] H. T. Gorji, R. Kardan, and N. Rezagholizadeh, "Analysis of blood gene expression data toward early detection of alzheimer's disease," *medRxiv,* p. 2021.07. 26.21261147, 2021.

[8] V. Atashi, R. Kardan, H. T. Gorji, and Y. H. Lim, "Comparative Study of Deep Learning LSTM and 1D-CNN Models for Real-time Flood Prediction in Red River of the North, USA," in *2023 IEEE International Conference on Electro Information Technology (eIT)*, 2023: IEEE, pp. 022-028.

[9] Z. Ahmad, S. Rahim, M. Zubair, and J. Abdul-Ghafar, "Artificial intelligence (AI) in medicine, current applications and future role with special emphasis on its potential and promise in pathology: present and future impact, obstacles including costs and acceptance among pathologists, practical and philosophical considerations. A comprehensive review," *Diagnostic pathology,* vol. 16, pp. 1-16, 2021.

[10] F. Afza, M. Sharif, M. Mittal, M. A. Khan, and D. J. Hemanth, "A hierarchical three-step superpixels and deep learning framework for skin lesion classification," *Methods,* vol. 202, pp. 88-102, 2022.

[11] M. A. Al-Masni, D.-H. Kim, and T.-S. Kim, "Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification," *Computer methods and programs in biomedicine,* vol. 190, p. 105351, 2020.

[12] S. S. Roy, A. U. Haque, and J. Neubert, "Automatic diagnosis of melanoma from dermoscopic image using real-time object detection," in *2018 52nd Annual Conference on Information Sciences and Systems (CISS)*, 2018: IEEE, pp. 1-5.

[13] S. S. Roy, R. Kardan, and J. Neubert, "A Deep Learning-Based Model for Melanoma Detection in Both Dermoscopic and Digital Images," in *2024 IEEE International Conference on Electro Information Technology (eIT)*, 2024: IEEE, pp. 668-673.

[14] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific data,* vol. 5, no. 1, pp. 1-9, 2018.

[15] A. Kirillov *et al.*, "Segment anything," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 4015-4026.

[16] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117-2125.

[17] T. Chu, X. Li, H. V. Vo, R. M. Summers, and E. Sizikova, "Improving weakly supervised lesion segmentation using multi-task learning," in *Medical Imaging with Deep Learning*, 2021: PMLR, pp. 60-73.

[18] H. Basak, R. Kundu, and R. Sarkar, "MFSNet: A multi focus segmentation network for skin lesion segmentation," *Pattern Recognition,* vol. 128, p. 108673, 2022.

[19] L. Yang, C. Fan, H. Lin, and Y. Qiu, "Rema-Net: An efficient multi-attention convolutional neural network for rapid skin lesion segmentation," *Computers in Biology and Medicine,* vol. 159, p. 106952, 2023.

[20] A. Bibi *et al.*, "Skin lesion segmentation and classification using conventional and deep learning based framework," *Comput. Mater. Contin,* vol. 71, no. 2, pp. 2477-2495, 2022.

[21] M. Karri, C. S. R. Annavarapu, and U. R. Acharya, "Skin lesion segmentation using two-phase cross-domain transfer learning framework," *Computer Methods and Programs in Biomedicine,* vol. 231, p. 107408, 2023.

[22] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. Marcal, and J. Rozeira, "PH 2-A dermoscopic image database for research and benchmarking," in *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, 2013: IEEE, pp. 5437-5440.

[23] I. Giotis, N. Molders, S. Land, M. Biehl, M. F. Jonkman, and N. Petkov, "MED-NODE: a computer-assisted melanoma diagnosis system using non-dermoscopic images," *Expert systems with applications,* vol. 42, no. 19, pp. 6578-6585, 2015.