# Highly Accurate Skin Cancer Diagnosis Using HMT-NET and Vision Transformer Models

**Sasikala N.[1]*, Vaka A. R.[2], Suguna R.[3], Kumar N.[4], Parvez M. M.[5], and Jamal K.[6]**

[1] Department of Electronics and Communication Engineering, Kamala Institute of Technology & Science, Singapur, Huzurabad, Karimnagar, Telangana 505468, India

[2] Department of Computer Science and Engineering, Lendi Institute of Engineering and Technology, Vizianagaram 535005, India

[3] Department of Computer Applications, Nandha Engineering College, Erode 638052, Tamilnadu, India

[4] Department of CSE, Amity University Jharkhand, Ranchi, Jharkhand 835303, India

[5] Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Green Fileds, Vaddeswaram, Guntur District, Andhra Pradesh 522502, India

[6] Department of Electronics and Communication Engineering, Gokaraju Rangaraju Institute of Engineering and Technology Bachupally, Kukatpally, Hyderabad, Telangana 500118, India

*e-mail: kalapraveen.sasi@gmail.com

**Abstract.** Skin cancer, especially melanoma, is one of the most aggressive and fatal cancers, with rising global incidence driven largely by ultraviolet exposure. Early detection is critical, as visual similarities among lesions complicate diagnosis. Clinical methods include self-examination, dermoscopic, and biopsy, while computational approaches range from traditional Computer-aided design systems to deep learning models like convolutional neural network and Vision Transformers. Although these models enhance accuracy, they face limitations such as high data requirements, limited interpretability, and inconsistent image quality, emphasizing the need for scalable and explainable diagnostic systems in clinical settings. A Hybrid Multi-layer Transformer Network (HMT-NET) and Vision Transformer (ViT)-based model is proposed for accurate skin lesion segmentation and multiclass skin cancer classification. This study presents a hybrid HMT-NET and ViT framework for accurate skin lesion segmentation and classification. Evaluated on HAM10000 and ISIC2019 datasets, the model achieved high Dice score and Jaccard index and 98.75% classification accuracy, demonstrating superior performance and reliability for integration into automated dermatological diagnostic systems.

**Keywords:** skin cancer; melanoma; early detection; dermoscopic; biopsy; self-examination; computer-aided diagnosis; convolutional neural network; vision transformer; HMT-NET; image segmentation; lesion classification; deep learning; medical image analysis; explainable AI; scalable diagnostic systems.

Paper #9345 received 28 Aug 2025; revised manuscript received 27 Oct 2025; accepted for publication 27 Oct 2025; published online 29 Dec 2025. doi: 10.18287/JBPE25.11.040308.

## 1 Introduction

Skin cancer, driven primarily by ultraviolet (UV) exposure, represents a major global health concern with rapidly rising incidence rates [1]. Melanoma, though less common, causes the majority of skin cancer deaths due to its aggressive nature. Early and accurate diagnosis is crucial but remains challenging because of the visual similarity between benign and malignant lesions [2, 3]. Traditional methods such as dermoscopy and biopsy, while effective, are subjective and unsuitable for large-scale screenings [4]. Computational approaches, particularly deep learning-based systems, have transformed dermatological diagnostics by automating lesion segmentation and classification [5, 6].

Early computer-aided diagnosis (CAD) systems relied on handcrafted features like color, texture, and asymmetry, analyzed through machine learning algorithms such as support vector machine (SVM) and *k*-nearest neighbors (KNN) [7, 8]. However, these models suffered from limited generalization and sensitivity to lighting conditions. The advent of convolutional neural networks (CNNs) and attention-based architectures significantly improved feature representation and classification accuracy [9, 10]. Hybrid models integrating CNNs and Transformers [11−13], such as BAT [14], LCAUnet [15], and Att-SwinU-Net [16], have achieved substantial gains in lesion segmentation by combining local detail extraction and global context modeling [17−19]. Vision Transformers (ViT) [20, 21] further enhance diagnostic performance by processing images as patch sequences, enabling the capture of long-range spatial dependencies crucial for skin lesion recognition [22].

Recent studies highlight the strength of combining CNNs and Transformers [23], as in HMT-Net [24] and DermViT, which demonstrated notable improvements across datasets like ISIC2019 and HAM10000. These approaches have motivated the current study's design of a hybrid HMT-NET and ViT framework for accurate, interpretable, and efficient skin cancer diagnosis.

## 2 Proposed Method

To address the challenges of accurate skin lesion segmentation and classification, a hybrid deep learning framework is proposed. This approach integrates a Transformer-enhanced segmentation network (HMT-NET) with a ViT-based classifier, aiming to enhance diagnostic precision by isolating lesion regions before classification. This design minimizes background interference and ensures that the model focuses on clinically relevant features for reliable diagnosis.

Figure 1 illustrates the workflow of the proposed hybrid framework for automated skin cancer classification, integrating HMT-NET for segmentation and ViT for classification. The process begins with the HAM10000 dataset, which provides dermoscopic skin lesion images and their corresponding binary masks. Initially, all images undergo preprocessing steps − resizing, contrast enhancement, and normalization − to ensure consistency and enhance feature clarity. These processed images and masks are then used to train HMT-NET, a hybrid architecture that combines Transformer and Multi-Layer Perceptron (MLP)-based mechanisms to capture both global contextual and local boundary information for precise lesion segmentation. Following training, the HMT-NET operates in two distinct phases: segmentation and classification training. During segmentation, unseen test images are preprocessed and passed through the trained HMT-NET to produce binary lesion masks. These masks are applied to the original images to generate region-of-interest (ROI) patches, isolating lesions while excluding irrelevant background regions and artifacts.

In parallel, the classification training phase employs a ViT trained on dermoscopic images and their corresponding class labels [3]. Each image is first processed by the trained HMT-NET to emphasize lesion-specific regions before being fed into the ViT. Leveraging its global self-attention mechanism, ViT models spatial dependencies across image patches to obtain a rich and discriminative feature representation, leading to accurate multi-class classification. The ViT is trained to distinguish among eight skin lesion classes: actinic keratoses (AK), basal cell carcinoma (BCC), benign keratosis-like lesions (BKL), dermatofibroma (DF), melanoma (MEL), melanocytic nevi (NV), squamous cell carcinoma (SCC), and vascular lesions (VASC). During inference, each masked test image is passed through the trained ViT, which outputs the predicted lesion category. This integrated framework – combining segmentation precision with transformer-based classification – enhances diagnostic accuracy by suppressing irrelevant background data and focusing on lesion-specific features, demonstrating strong potential for real-world clinical deployment.

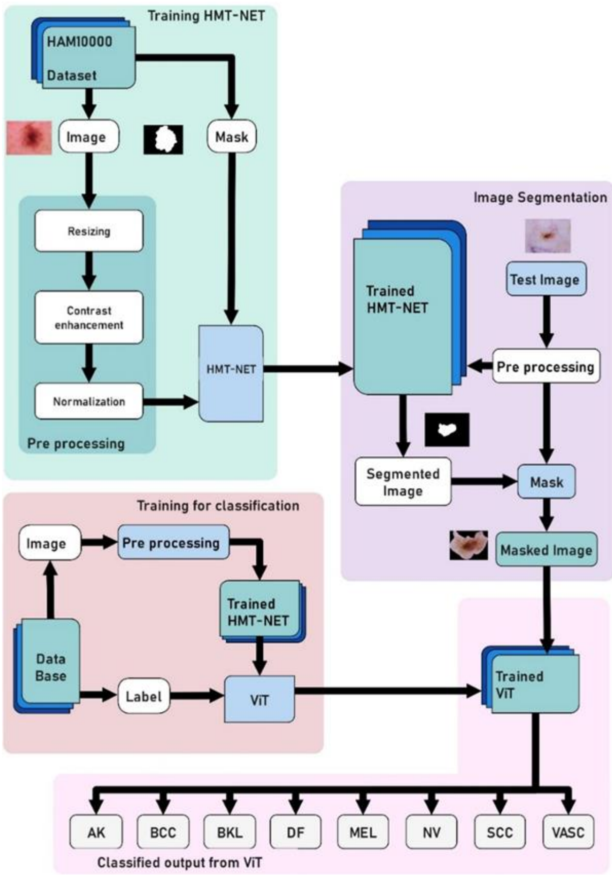



Fig. 1 Block diagram of the proposed skin cancer classification.

### *2.1 Preprocessing*

The preprocessing stage includes three key steps – resizing, contrast enhancement, and normalization –

designed to standardize and optimize Dermoscopic images for model input.

Resizing ensures that all images conform to a fixed resolution compatible with the network architecture, minimizing variations caused by differing acquisition settings [25]. Contrast enhancement improves lesion visibility and boundary clarity, addressing challenges such as low contrast or uneven lighting common in dermoscopic images. Finally, normalization scales pixel values (typically within [0, 1] or standardized to zero mean and unit variance), reducing illumination variability and stabilizing model training. Together, these steps improve image consistency, highlight clinically relevant features, and enhance convergence stability for the subsequent deep learning stages.

## *2.2 HMT-NET*

The HMT-NET architecture performs precise lesion segmentation by combining convolutional encoding, transformer-based global attention, and MLP-driven refinement. As shown in Fig. 2, it comprises three main components: a multi-stage encoder ($G_1$–$G_5$), a dual-attention pathway (Convolutional Transformer (CTrans) and Token-Mixing Multi-Layer Perceptron (ToK-MLP)), and a decoder that reconstructs the final segmentation mask [26].

The encoder progressively extracts hierarchical features – ranging from shallow textures to deep semantic representations. The CTrans module applies multi-head self-attention to capture long-range dependencies, while the ToK-MLP module enhances edge precision and local boundary details [6]. The decoder then fuses these refined features through up-sampling and skip connections to produce an accurate binary lesion mask. This hybrid design allows HMT-NET to effectively integrate global contextual understanding with local structural detail, resulting in highly accurate segmentation of complex skin lesions in Dermoscopic imagery.

## *2.3 ViT*

The ViT is a deep learning architecture that adapts the Transformer model, originally developed for natural language processing, to image classification tasks. Unlike CNNs that capture local spatial hierarchies, ViT represents an image as a sequence of fixed-size patches, enabling the model to learn global contextual relationships across the entire image early in the processing pipeline.

Figure 3 illustrates the workflow of ViT used for skin lesion classification. The input dermoscopic image is divided into non-overlapping patches, each flattened and projected into an embedding vector. Positional encodings are added to preserve spatial information, and a special classification token ([CLS]) is prepended to aggregate image-level features. The sequence is then processed through multiple Transformer layers consisting of multi-head self-attention (MHSA) and feedforward MLP blocks, each followed by normalization and residual connections.

For classification, the final representation of the [CLS] token is passed through fully connected layers and a SoftMax activation to predict lesion categories. The model is trained to classify eight skin lesion types: AK, BCC, BKL, DF, MEL, NV, SCC, and VASC.

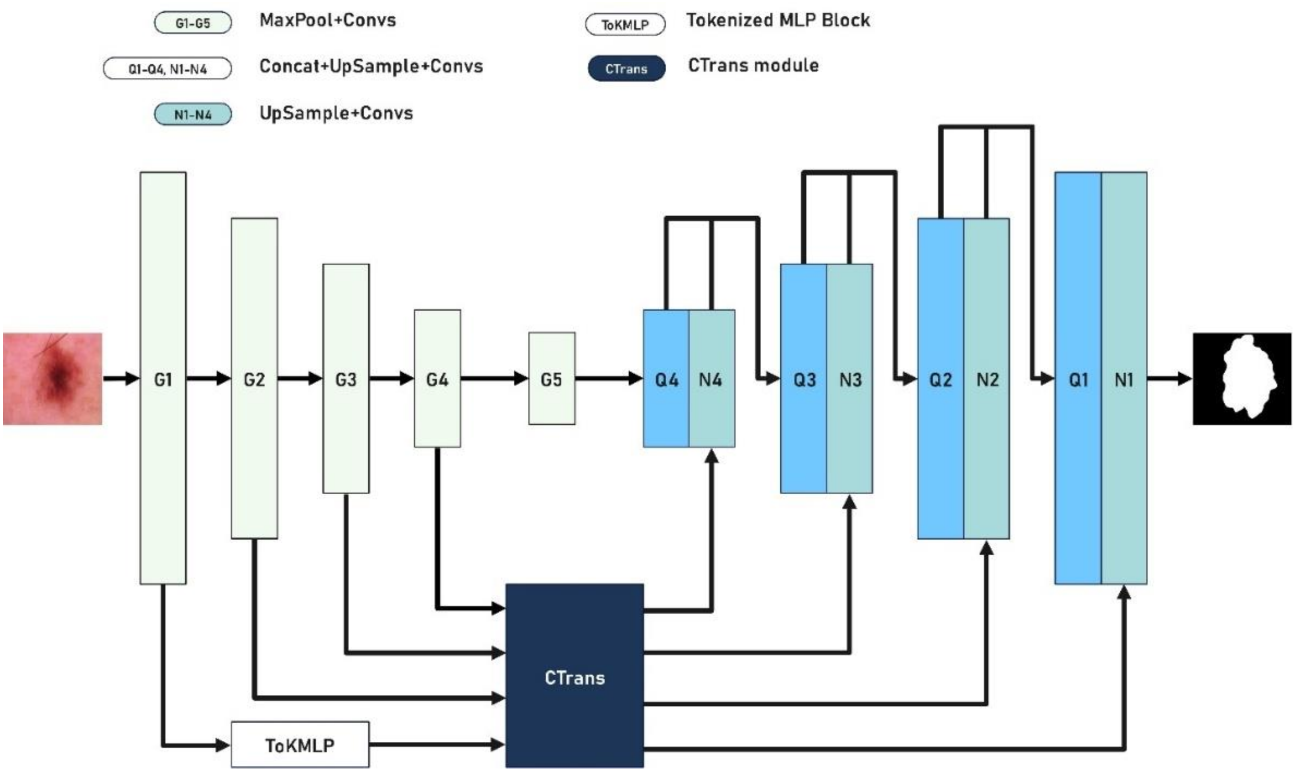



Fig. 2 Block diagram of the HMT-NET working.

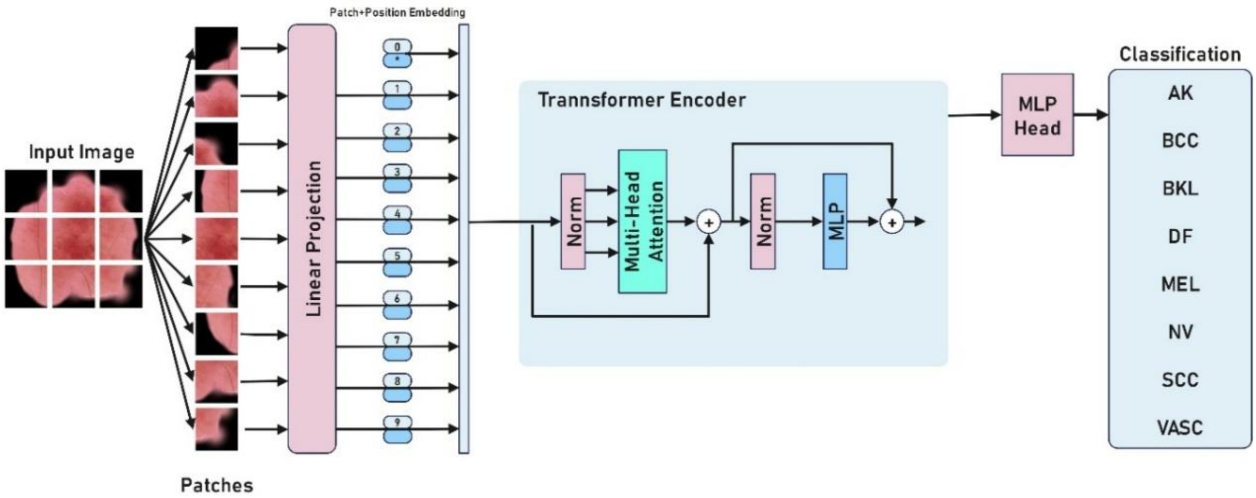



Fig. 3 Block diagram of Vision Transformer working.

## 3 Results and Discussion

This section summarizes the experimental results of the proposed HMT-NET + ViT framework for skin lesion segmentation and multi-class classification. Experiments were conducted on the HAM10000 and ISIC2019 datasets using a GPU-enabled Google Colab environment. Evaluation metrics included Dice Coefficient, Intersection over Union (IoU), accuracy, precision, recall, and $F_1$-score, computed per image and averaged across the datasets for fair comparison.

### *3.1 Dataset*

All images underwent preprocessing for resizing, contrast enhancement, and normalization. The HAM10000 dataset (seven lesion classes) was used for segmentation, while the ISIC2019 dataset (eight classes: AK, BCC, BKL, DF, MEL, NV, SCC, VASC) was employed for classification. Macro-averaging was applied to mitigate class imbalance.

### *HAM10000 Segmentation Performance*

The HMT-NET achieved a Dice score of 91.85% and an IoU of 88.42%, surpassing baseline architectures such as U-Net, R2U-Net, and FAC-Net. This improvement arises from the CTrans attention for long-range feature capture and the tokenized MLP for precise boundary refinement. Qualitative outputs (Fig. 4) show accurate lesion delineation even under challenging conditions such as hair occlusion and low contrast. Table 1 shows the comparative performance of the proposed method HMT-NET.

### *ISIC2019 Classification*

Using segmented ROIs from HMT-NET, the ViT-based classifier achieved 89.72% top-1 accuracy and a macro $F_1$-score of 88.35%, outperforming ResNet-50, EfficientNet, and DenseNet. High $F_1$-scores (> 90%) for NV and MEL classes confirm strong discrimination between benign and malignant lesions. The segmentation-guided input significantly reduced false positives by eliminating irrelevant background, while ViT's self-attention effectively captured subtle inter-class variations. The confusion matrix (Fig. 5) illustrates consistently high class-wise accuracy, especially for MEL and SCC, validating the model's reliability in real-world diagnostic settings.

### *3.2 Performance Metrics*

To assess the effectiveness of the proposed hybrid HMT-NET and ViT framework, a set of standard performance metrics was employed, focusing on both segmentation quality (HAM10000 dataset) and classification accuracy (ISIC2019 dataset).

### *Accuracy*

Accuracy measures the proportion of correctly classified pixels (for segmentation) or samples (for classification) among the total.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}. \quad (1)$$

### *Dice Coefficient (DSC)*

Also known as the $F_1$-score in segmentation, Dice measures the overlap between predicted and ground truth masks.

$$DSC = \frac{2TP}{FP + 2TP + FN}. \quad (2)$$

### *Jaccard Index*

Also known as IoU, this metric evaluates the intersection divided by the union of predicted and actual segmentations.

$$Jaccard = \frac{TP}{TP+FP+FN}. \quad (3)$$

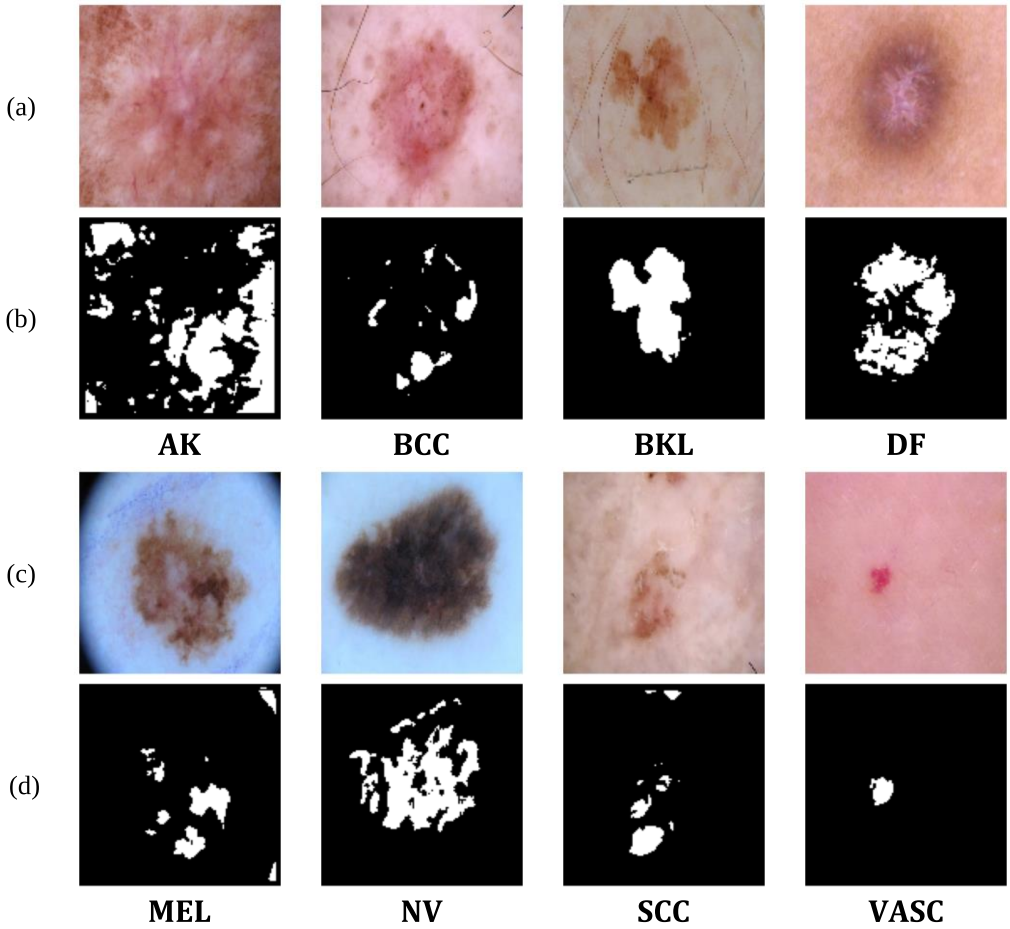



Fig. 4 Segmented output of classified image: (a) input image of HAM10000, (b) segmented image of HAM10000, (c) input image of ISIC2019, (d) segmented image of ISIC2019.

Table 1 Comparative performance of the proposed method HMT-NET.

| Model | Accuracy (%) | Dice Coefficient (%) | Jaccard Index (%) | Sensitivity (%) | Specificity (%) |
|---|---|---|---|---|---|
| Detectron2 (Mask R-CNN) | 98.40 | 94.00 | 86.90 | – | – |
| Detectron2 (Mask R-CNN) | 99.80 | 98.50 | 94.20 | – | – |
| Detectron2 (Mask R-CNN) | 98.90 | 96.10 | 91.50 | – | – |
| Ultra-Light VM-U Net | 96.46 | 90.91 | 83.34 | 90.53 | 97.9 |
| Ultralight VM-U Net | 95.58 | 89.40 | 80.56 | 86.80 | 97.81 |
| Ultralight VM-U Net | 95.21 | 92.65 | 86.31 | 93.45 | 96.06 |
| CNN + Classifier | 98.34 | 97.15 | – | 97.25 | 99.13 |
| UCM-NetV2 | 96.82 | 91.54 | – | 90.85 | 98.22 |
| UCM-NetV2 | 96.10 | 90.74 | – | 90.43 | 97.77 |
| ESC-UNET | 98.25 | 95.15 | 88.41 | 95.67 | 99.24 |
| ESC-UNET | 98.30 | 93.48 | 83.88 | 94.17 | 99.32 |
| ESC-UNET | 97.89 | 92.53 | 81.67 | 93.63 | 99.03 |
| CCT-Net | – | 93.21 | 87.70 | – | – |
| CCT-Net | – | 89.19 | 82.05 | – | – |
| CCT-Net | – | 92.72 | 86.81 | – | – |
| CCT-Net | – | 96.21 | 92.79 | – | – |

Table 1 Cont.

| | | | | | |
|---|---|---|---|---|---|
| Ensemble-A | 94.10 | 87.10 | 79.30 | 89.90 | 95.00 |
| Ensemble-S | 93.80 | 90.70 | 83.90 | 93.20 | 92.90 |
| O-Net | 94.71 | 87.04 | 80.36 | 89.70 | 96.30 |
| O-Net | 95.14 | 92.12 | 86.15 | 89.23 | 96.75 |
| YOLOv4+ActiveContour | 93.90 | 92.00 | 96.00 | 94.00 | 95.20 |
| YOLOv4+ActiveContour | 95.00 | 89.00 | 98.00 | 93.20 | 93.45 |
| Used Method | 98.00 | 91.00 | 98.00 | 93.60 | 93.88 |

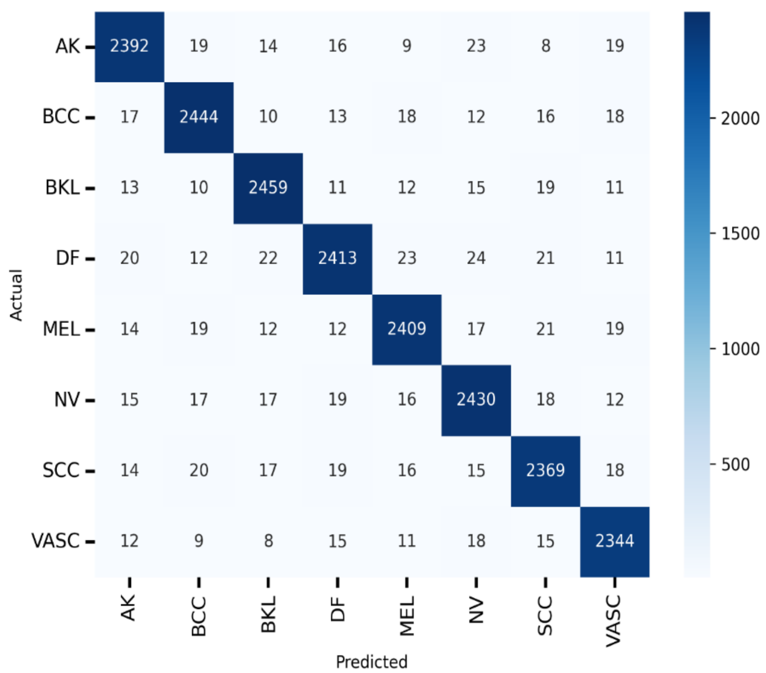



Fig. 5 Confusion matrix of classification.

Table 2 Comparative performance of the proposed method ViT.

| **Model Used** | **Accuracy** | **Precision** | **Recall** | ***$F_1$*-Score** |
|---|---|---|---|---|
| IncepX + Ensemble (SVM, RF, KNN) | 97.32% | 96.80% | 96.59% | 96.60% |
| Class-weighted RL + ResNet50 | 91.62% | 90.13% | 91.72% | 90.91% |
| YOLOv8 + Vision Transformer (ViT) | 98.60% | 98.10% | 98.40% | 98.20% |
| Generative AI + Reset Variant | 95.40% | 94.30% | 94.90% | 94.60% |
| Interpretable CNN + CBAM (Attention) | 93.28% | 92.91% | 93.12% | 93.00% |
| Generative AI + Deep CNN | 95.40% | 94.30% | 94.90% | 94.60% |
| Enhanced CNN + Grad-CAM Attention | 93.70% | 93.10% | 92.80% | 92.90% |
| Shuffle Net + Non-local Attention | 96.12% | 95.86% | 95.90% | 95.88% |
| Multi-Model Attentional Fusion Ensemble | 98.40% | 97.90% | 98.10% | 98.00% |
| Hybrid CNN + SVM (post-segmentation) | 92.00% | – | – | – |
| SVM, RF on Gene Data | 91.20% | 90.60% | 90.90% | 90.70% |
| FCN-Based Dense Net Framework | 94.31% | 94.00% | 94.35% | 94.12% |
| Used Method | 98.75% | 96.37% | 98.66% | 98.90% |

### *Recall (Sensitivity)*

Specificity assesses the proportion of actual negatives correctly identified.

$$Sensitivity = \frac{TP}{TP+FN}. \quad (4)$$

### *Specificity*

Specificity assesses the proportion of actual negatives correctly identified.

$$Specificity = \frac{TN}{TN+FP}. \quad (5)$$

Table 2 summarizes comparisons with state-of-the-art frameworks such as YOLOv8 + ViT, Multi-Model Attentional Fusion, and IncepX + SVM/RF/KNN.

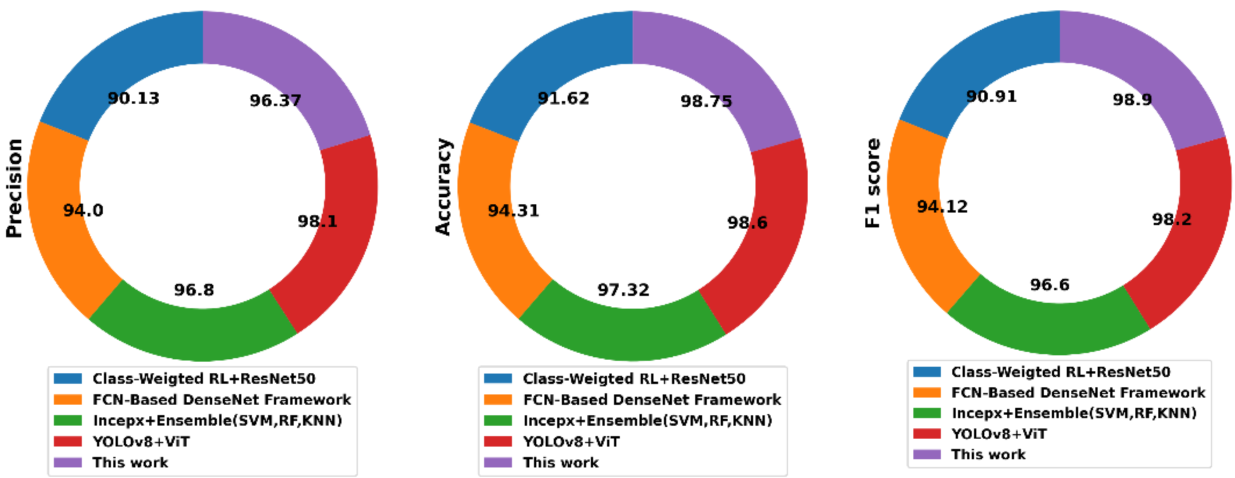
While existing attention-based ensembles reach ~98% accuracy, the proposed HMT-NET + ViT achieved the highest overall performance − 98.75% accuracy, 96.37% precision, 98.66% recall, and 98.90% $F_1$-score − demonstrating superior diagnostic precision and generalization.

Class-wise metrics (Table 3) show uniformly strong results across all categories, with near-perfect accuracy for NV, BCC, and SCC, and significant improvement in MEL classification ($F_1$ = 0.982). This balanced performance highlights the framework's ability to handle data imbalance and complex lesion textures effectively.

Across both datasets, the hybrid HMT-NET + ViT pipeline consistently outperformed state-of-the-art methods in segmentation and classification. Its integration of transformer-based global attention with MLP-driven local refinement ensures fine boundary detection and robust contextual understanding. With high accuracy, balanced sensitivity/specificity, and efficient computation, the framework demonstrates strong potential for real-time, clinically interpretable skin cancer diagnosis. Fig. 6 shows the performance of the proposed method.

Table 3 Classification performance by classes.

| Class | Accuracy | Precision | Recall | $F_1$-Score |
|---|---|---|---|---|
| AK | 0.987 | 0.961 | 0.984 | 0.987 |
| BCC | 0.988 | 0.965 | 0.987 | 0.989 |
| BKL | 0.986 | 0.957 | 0.988 | 0.986 |
| DF | 0.988 | 0.962 | 0.986 | 0.987 |
| MEL | 0.982 | 0.942 | 0.976 | 0.982 |
| NV | 0.992 | 0.981 | 0.992 | 0.991 |
| SCC | 0.989 | 0.966 | 0.986 | 0.988 |
| VASC | 0.991 | 0.969 | 0.987 | 0.989 |




(a) (b) (c)

Fig. 6 Performance of the proposed method. (a) Performance of precision, (b) performance of accuracy, (c) performance of $F_1$-score.

## 4 Conclusion

This study presents a hybrid diagnostic framework integrating the HMT-NET segmentation model with a Vision Transformer (ViT) classifier for effective skin cancer detection. The proposed pipeline was evaluated on the HAM10000 and ISIC2019 datasets using GPU-enabled platforms. Experimental results demonstrated that HMT-NET achieved a Dice coefficient of 91.85% and an IoU of 88.42%, outperforming established models such as U-Net, R2U-Net, and FAC-Net. These improvements are attributed to the use of tokenized MLP and attention-based components, which enhanced lesion boundary detection even under challenging conditions like hair occlusion and low contrast. For classification, the ViT model achieved a top-1 accuracy of 89.72% and a macro averaged $F_1$-score of 88.35%. When compared to models such as YOLOv8+ViT and IncepX-Ensemble, our method surpassed existing solutions with a final classification accuracy of 98.75% and $F_1$-score of 98.90%. Class-wise results further confirmed consistent performance, particularly for high-risk lesions such as NV and SCC. These findings demonstrate the framework's robustness, scalability, and clinical relevance. Overall, the proposed method shows strong potential for integration into real-time, automated dermatological diagnostic systems, offering improved accuracy, explain ability, and efficiency for multi-class skin lesion detection in practical healthcare applications.

## Disclosures

All authors declare that there is no conflict of interest in this paper.

## References

1. A. C. Green, C. M. Olsen, and D. J. Hunter, Skin Cancer, in Textbook of cancer epidemiology, H.-O. Adami, D. Hunter, D. Trichopoulos (Ed.), 2nd ed., Oxford University Press, New York (2018). ISBN: 978-0-19-531117-4
2. D. Moturi, R. K. Surapaneni, and V. S. G. Avanigadda, "Developing an efficient method for melanoma detection using CNN techniques," Journal of the Egyptian National Cancer Institute 36(1), 6 (2024).
3. H. AbuAlkebash, R. A. A. Saleh, and H. M. Ertunç, "Automated explainable deep learning framework for multiclass skin cancer detection and classification using hybrid YOLOv8 and vision transformer (ViT)," Biomedical Signal Processing and Control 108, 107934 (2025).
4. N. Melarkode, K. Srinivasan, S. M. Qaisar, and P. Plawiak, "AI-powered diagnosis of skin cancer: a contemporary review, open challenges and future research directions," Cancers 15(4), 1183 (2023).
5. J. Wang, L. Wei, L. Wang, Q. Zhou, L. Zhu, and J. Qin, "Boundary-aware transformers for skin lesion segmentation," in Medical Image Computing and Computer Assisted Intervention – MICCAI 2021, M. De Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert (Eds.), 12901, Springer, Cham, 206–216 (2021).
6. G. Wang, Q. Ma, Y. Li, K. Mao, L. Xu, and Y. Zhao, "A skin lesion segmentation network with edge and body fusion," Applied Soft Computing 170, 112683 (2025).
7. M. A. Arshed, S. Mumtaz, M. Ibrahim, S. Ahmed, M. Tahir, and M. Shafi, "Multi-class skin cancer classification using vision transformer networks and convolutional neural network-based pre-trained models," Information 14(7), 415 (2023).
8. I. Ahmed, B. Bushon Routh, Md. S. Rahman Kohinoor, S. Sakib, M. Mahfuzur Rahman, and F. Azzedin, "Multi-model attentional fusion ensemble for accurate skin cancer classification," IEEE Access 12, 181009–181024 (2024).
9. E. K. Aghdam, R. Azad, M. Zarvani, and D. Merhof, "Attention swin U-Net: Cross-contextual attention mechanism for skin lesion segmentation," in IEEE 20th International Symposium on Biomedical Imaging, IEEE, 1–5 (2023).
10. S. Yang, L. Wang, "HMT-Net: Transformer and MLP hybrid encoder for skin disease segmentation," Sensors 23(6), 3067 (2023)
11. X. Zhang, Y. Liu, G. Ouyang, W. Chen, A. Xu, T. Hara, X. Zhou, and D. Wu, "DermViT: Diagnosis-guided vision transformer for robust and efficient skin lesion classification," Bioengineering 12(4), 421 (2025).
12. K. Rezaee, H. G. Zadeh, "Self-attention transformer unit-based deep learning framework for skin lesions classification in smart healthcare," Discover Applied Sciences 6(1), 3 (2024).
13. I. Pacal, M. Alaftekin, and F. D. Zengul, "Enhancing Skin cancer diagnosis using swin transformer with hybrid shifted window-based multi-head self-attention and SwiGLU-based MLP," Journal of Imaging Informatics in Medicine 37(6), 3174–3192 (2024).
14. Y. Mo, P. Zuo, Q. Zhou, Z. Mo, Y. Fan, S. Zhang, and B. Kang, "PWLT: Pyramid window-based lightweight transformer for image classification," Computers and Electrical Engineering 116, 109209 (2024).
15. S. Aladhadh, M. Alsanea, M. Aloraini, T. Khan, S. Habib, and M. Islam, "An effective skin cancer classification mechanism via medical vision transformer," Sensors 22(11), 4008 (2022).
16. G. M. S. Himel, Md. M. Islam, Kh. A. Al-Aff, S. I. Karim, and Md. K. U. Sikder, "Skin cancer segmentation and classification using vision transformer for automatic analysis in dermatoscopy-based noninvasive digital system," International Journal of Biomedical Imaging 2024, 1–18 (2024).
17. R. Wu, Y. Liu, G. Ning, P. Liang, and Q. Chang, "UltraLight VM-UNet: Parallel Vision Mamba significantly reduces parameters for skin lesion segmentation," Patterns 6(11), 101298 (2025).

18. A. Alrabai, A. Echtioui, and F. Kallel, "Explainable deep learning approaches for skin cancer diagnosis," Network Modeling and Analysis in Health Informatics and Bioinformatics Journal 14(1), 57 (2025).
19. H. Amjad, N. Asif, H. Elahi, U. Shahbaz Khan, H. Akbar, A. R. Ansari, and R. Nawaz, "Precision segmentation and binary masking of skin lesions in automated dermatological diagnostics using detectron2," IEEE Access 12, 187696–187708 (2024).
20. S. Chatterjee, J.-M. Gil, and Y.-C. Byun, "Early detection of multiclass skin lesions using transfer learning-based IncepX-ensemble model," IEEE Access 12, 113677–113693 (2024).
21. C. Yuan, D. Zhao, and S. S. Agaian, "UCM-NetV2: An Efficient Deep Learning Model for Skin Lesion Segmentation," Journal of Economy and Technology 3, 251–263 (2025).
22. Z. Ji, X. Wang, C. Liu, Z. Wang, N. Yuan, and I. Ganchev, "EFAM-Net: A multi-class skin lesion classification model utilizing enhanced feature fusion and attention mechanisms," IEEE Access 12, 143029–143041 (2024).
23. A. Jimi, N. Zrira, O. Guendoul, I. Benmiloud, H. A. Khan, and S. Nawaz, "ESC-UNET: A hybrid CNN and swin transformers for skin lesion segmentation," Intelligence-Based Medicine 12, 100257 (2025).
24. R. Karthik, R. Menaka, S. Atre, J. Cho, and S. Veerappampalayam Easwaramoorthy, "A hybrid deep learning approach for skin cancer classification using swin transformer and dense group shuffle non-local attention network," IEEE Access 12, 158040–158051 (2024).
25. A. D. Khalaf, H. Hamdan, A. Abdul Halin, and N. Manshor, "Segmentation and classification of skin cancer diseases based on deep learning: Challenges and future directions," IEEE Access 13, 90163–90184 (2025).
26. V. Kumar, D. L. Shanthi, T. R. Babu, N. Kumar, and R. K. Godi, "Advanced Skin Lesion Segmentation Using Adaptive Contextual GLCM," Egyptian Informatics Journal 30, 100706 (2025).