An Enhanced Skin Cancer Detection Method Utilizing TriBlendNet and Deepdilated Focus U-Net

D. Manju
Department of CSE-(CyS, DS) and AI&DS, Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India
nuthana525@gmail.com (corresponding author)

K. Kishore Kumar
Department of Electronics and Communication Engineering, Faculty of Science & Technology (IcfaiTech), ICFAI Foundation for Higher Education, Hyderabad, India
kishorekamarajugadda@gmail.com

Movva Pavani
Department of ECE, Nalla Malla Reddy Engineering College, Divyanagar, Hyderabad, India
drmovvapavani@gmail.com

N. V. S. Pavan Kumar
Department of Computer Science and Engineering, Koneru Lakshmiah Education Foundation, India
nvspavankumar@kluniversity.in

V. S. N. Murthy
Department of Information Technology, Shri Vishnu Engineering College for Women, Bhimavaram, India
vsn.murthy87@gmail.com

Rajesh Kumar Verma
Department of CSE, CMR College of Engineering, Hyderabad, India
rajeshverma.hyd10@gmail.com

Padmini Debbarma
ICFAI University, Hyderabad, India
debbarmapadmini.scholar24@ifheindia.org

M. Koteswara Rao
Department of Information Technology, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana, India
mailtomkrao@gmail.com

Anand Kumar Saraswathi Rathod
Department of CSE-(CyS,DS) and AI&DS, Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India
sr_ak@yahoo.com

Bh. Krishna Mohan
Department of CSE (AI–ML), RVR & JC College of Engineering, Guntur, Andhra Pradesh, India
bkm@rvrjc.ac.in

Received: 22 August 2025 | Revised: 5 October 2025 and 21 October 2025 | Accepted: 3 November 2025

 DOI: https://doi.org/10.48084/etasr.14192

ABSTRACT

Skin cancer remains a major global health concern, demanding advanced methods for accurate and early identification. This work presents an integrated framework that employs a Generative Adversarial Network (GAN) for effective data augmentation and a novel Deep Dilated-Focus U-Net enhanced with attention mechanisms for precise lesion segmentation. For classification, a hybrid model named TriBlendNet is proposed, combining the advantages of the SqueezeNet and DenseNet121 architectures. Using the SIIM-ISIC 2019 dataset, the proposed system outperforms existing models such as SqueezeNet, DenseNet121, ResNet50, and VGG16. The TriBlendNet model achieved an outstanding accuracy of 98.59%, along with high precision, recall, and specificity, showcasing its strong capability for reliable and efficient automated skin cancer detection.

Keywords-skin cancer; deep dilated-focus U-Net; Neural Architecture Search (NAS); TriBlendNet

I. INTRODUCTION

Skin cancer occurs when skin cells undergo malignant changes, mainly due to prolonged exposure to Ultraviolet (UV) radiation [1, 2]. The main types are Basal Cell Carcinoma (BCC), Squamous Cell Carcinoma (SCC), and melanoma, the latter being the most aggressive, though less common. Melanoma develops in melanocytes, appearing as irregular, pigmented lesions with uneven borders. To address diagnostic challenges, this study proposes TriBlendNet as an automated detection system that employs an Attention-Enhanced U-Net with hybrid hyperparameter optimization for precise lesion segmentation and classification, aiming to reduce false positives/negatives and supporting clinicians with accurate and reliable diagnoses. The main contributions of this work are as follows:

- Deep Dilated-Focus U-Net for segmentation: This architecture is a variant of U-Net, augmented with dilated convolutions and attention mechanisms inspired by the Convolutional Block Attention Module (CBAM), which enables very precise Region of Interest (ROI) segmentation.
- TriBlendNet classification model: This hybrid model leverages the best practices of both the DenseNet121 and SqueezeNet architectures to create a superior and computationally efficient skin cancer detector.

In [1], an optimized CNN enhanced feature extraction and classification for dermoscopic images. In [2], a hybrid model combined fuzzy logic with deep learning, enabling better handling of uncertainties in lesion patterns. SCDet [3] improves lesion localization and classification through advanced deep learning architectures, providing a reliable tool for clinical applications. DSCC_Net [4] is a multi-class deep learning model capable of categorizing various lesion types, facilitating automation in diagnostic processes. In [5], Vision Transformers (ViTs) were combined with feature selection techniques, capturing global image context and improving classification accuracy. In [6], super-resolution GANs were combined with custom CNNs, enhancing image quality and feature extraction for superior performance. In [7], it was demonstrated that analyzing multiple dermoscopic images per lesion improves melanoma classification, emphasizing the benefit of multi-view analysis. In [8], hybrid deep learning models improved sensitivity and specificity in clinical settings. Widely used benchmark datasets, such as SIIM-ISIC [9] and HAM10000 [10], provide standardized platforms for training and evaluating these models, with [11] reporting improved classification across multiple lesion categories. In [12], lesion segmentation was enhanced through boundary-aware networks combined with convolutional and transformer-based models. Modern architectures, including EfficientNetV2 [13], "A ConvNet" [14], ViTs [15], and GANs [16], have further improved efficiency, accuracy, and image quality. In [17], CNN architectures were optimized with checkpoint strategies, while the study in [18] demonstrated the effectiveness of transformer-based systems in segmentation and classification. In [19], CNNs and transformers were compared, highlighting the advantages and limitations of each approach for automated skin disease analysis. Collectively, these studies illustrate a clear progression towards hybrid and transformer-based models, leveraging high-quality datasets and advanced architectures to enhance diagnostic performance.

II. PROPOSED METHOD

The proposed method combines GAN-based augmentation, adaptive batch normalization, and Gaussian blur for robust preprocessing. For segmentation, a dilated U-Net [20] with CBAM attention is employed [21], followed by classification through the hybrid TriBlendNet model. Figure 1 illustrates the overall pipeline workflow. The process begins with an input image of a skin lesion, first augmented using a GAN [14] to expand the training dataset. The image is then preprocessed with Gaussian blur and adaptive batch normalization to reduce noise and standardize the data. Key features are identified, and a GazFOX optimization step fine-tunes parameters such as batch size. The core of the system is the TriBlendNet model, a custom architecture that integrates efficient components from SqueezeNet and DenseNet. The processed information is used to classify the lesion, producing a definitive output of either Skin Cancer or Non-Cancer. The utilized dataset comes from

the SIIM-ISIC 2019 collection [9], comprising 25,331 dermoscopic images spanning eight categories: Melanoma, Seborrheic Keratosis, Dermatofibroma, Vascular Lesion, Basal Cell Carcinoma, Squamous Cell Carcinoma, and Actinic Keratosis, with 4,522 images being melanoma, which is the primary focus of this research. All images are resized to 512×512 pixels to maintain consistency and facilitate effective model training.

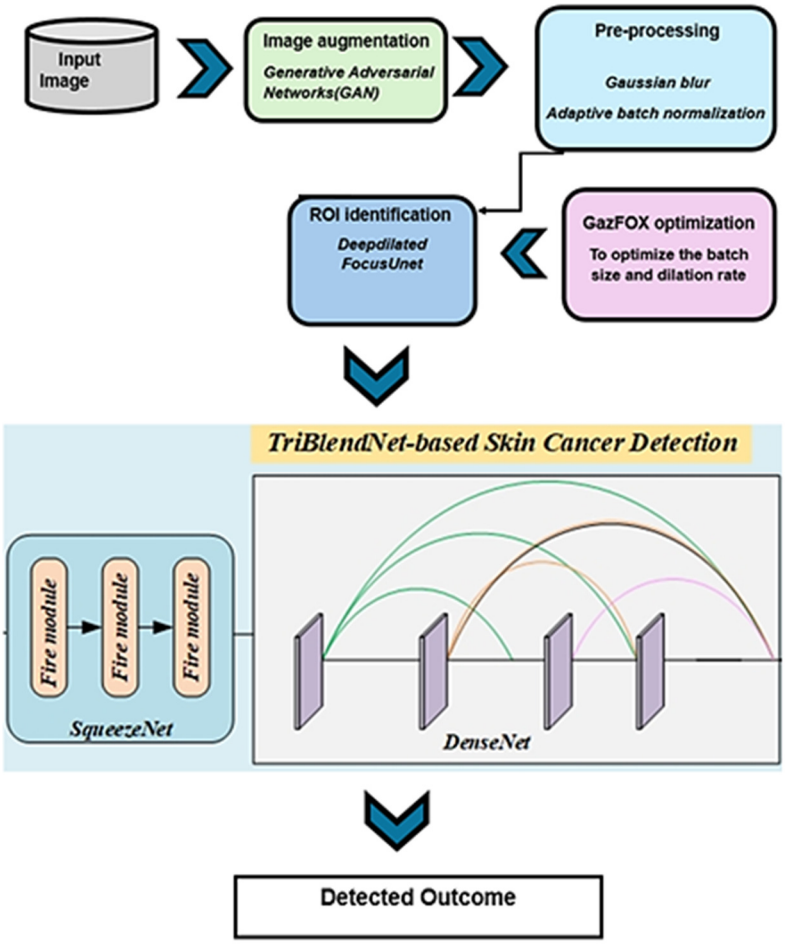

Fig. 1. The proposed framework.

Several metrics were employed to evaluate the model's diagnostic performance and clinical reliability:

- Accuracy is the ratio of correctly classified instances to the total number of cases.
- Precision is the proportion of predicted positive cases that are actually positive.
- Recall (Sensitivity) is the proportion of true positive cases correctly identified.
- F1-score is the harmonic mean of precision and recall, providing a balanced assessment.
- Specificity is the proportion of true negative cases correctly classified.
- Sensitivity measures the model's ability to correctly detect positive cases.

A. *Image Augmentation*

GANs are capable of generating synthetic images that closely resemble actual skin lesion images in the context of image augmentation. The discriminator loss function [14] is given by:

$$\frac{max}{D} V(G,D) = E_{y\sim P_{data}(y)}[logD(y)] + E_{n\sim P_n(n)}[log\left(1 - D\big(G(n)\big)\right)] \quad (1)$$

The generator is trained to minimize $log\left(1 - D\big(G(n)\big)\right)$, creating images that are as close to the actual training images as possible. The generator loss function is expressed as

$$\frac{min}{G} V(G,D) = E_{y\sim P_n(n)}[log\,(1 - D\big(G(n)\big)] \quad (2)$$

The generator and discriminator are optimized using:

$$\frac{min}{G}\frac{max}{D} V(D,G) = E_{y\sim P_{data}(y)}[logD(y)] + E_{n\sim P_n(n)}\big[\log\big(1 - D\big(G(n)\big)\big]\big] \quad (3)$$

B. *Preprocessing*

In situations where it could be challenging to detect malignant areas, Gaussian blur aids in accurate identification. Adaptive batch normalization works in tandem with

normalizing to enhance dataset robustness and ensure that the images are consistently and effectively prepared for subsequent analysis stages. The 2D Gaussian function to blur images is

$$G(x,y) = \frac{1}{2\pi\sigma^2}.e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)} \quad (4)$$

Adaptive Batch Normalization (BN) is the implementation of domain-specific normalization for different domains. In layers of BN, the average (μ_j) and variance (σ_j^2) can be updated by applying the following equation to a batch of k simples for neuron j:

$$d = \mu - \mu_j \quad (5)$$

$$\mu_j \leftarrow \mu_j + \frac{dk}{n_j} \quad (6)$$

$$\sigma_j^2 \leftarrow \frac{\sigma_j^2 n_j}{n_j+k} + \frac{\sigma^2 k}{n_j+k} + \frac{d^2 n_j k}{(n_j+k)^2} \quad (7)$$

$$n_j \leftarrow n_j + k \quad (8)$$

where n_j is the stored statistic representing the number of samples for neuron j in previous iterations, and μ and σ^2 are the mean and variance of the current input batch for neuron j. The variance σ_j^2 and mean μ_j are initialized to one and zero, respectively.

C. Deepdilated Focus U-Net for ROI

The energy function of the proposed Deep Dilated-Focus U-Net is given by:

$$E = w(x)log(pk(x)(x)) \quad (9)$$

where p_k is the SoftMax function applied pixel-wise over the final feature map:

$$p_k = \frac{1\exp(a_k(x))}{\sum_{k'=1}^{K} 1\exp(a_k(x)')} \quad (10)$$

and the activation in channel k is indicated by a_k.

D. CBAM

The following equations define the procedure by which CBAM processes the input feature map F:

$$F' = M_c(F) \otimes F \quad (11)$$

$$F'' = M_s(F') \otimes F' \quad (12)$$

where F' represents the outcome of multiplying the feature map by the channel attention map, $\otimes$ indicates element multiplication, and F'' represents the refined output in its final form. For the channel attention module, two layers perform better than any other single layer. The following equation is used for the channel attention:

$$M_c(F) = \sigma\left(MLP\left(AvgPool(F)\right) + MLP\left(MaxPool(F)\right)\right) \quad (13)$$

$$M_c(F) = \sigma\left(W_1\left(W_0\left(F_{avg}^c\right)\right) + W_1\left(W_0\left(F_{max}^c\right)\right)\right) \quad (14)$$

where σ represents the sigmoid function, $W_1 \in R^{C\times C/r}$, and $W_0 \in R^{C/r\times C}$. It should be noted that W_0 comes after the ReLU activation function and W_0 and W_1 are MLP weights shared by the two inputs. The following equation describes how this is multiplied by the input feature map to generate the final result.

$$M_s(F) = \sigma\left(f^{7\times7}([AvgPool(F); MaxPool(F)]\right) \quad (15)$$

$$M_s(F) = \sigma\left(f^{7\times7}\left[F_{avg}^s; F_{max}^s\right]\right) \quad (16)$$

where $f^{7\times7}$ represents a convolution operation with a 7×7 filter size, and σ is a sigmoid function. To create a channel attention map, CAM makes use of the features' inter-channel relationships.

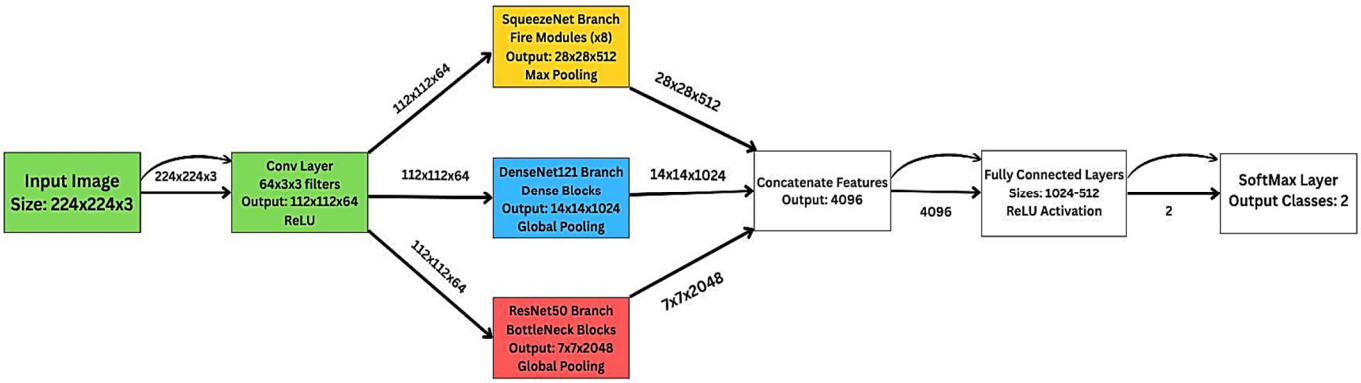

Fig. 2. Deep dilated focus U-Net.

E. TriBlend Net

Figure 2 illustrates the TriblendNet architecture. Its novelty is not only the combination but the orchestrated synergy where these architectures work in parallel to extract complementary feature types (textural, hierarchical, and semantic) that are all critical for dermatological diagnosis. A simple ensemble averages predictions, but TriBlendNet's design allows for a learned fusion of these feature maps before the final classification layer, creating a richer, more discriminative feature representation. TriBlendNet mainly combines existing models (SqueezeNet, DenseNet121, ResNet50).

The role of the SqueezeNet branch is not just parameter efficiency but to act as a high-frequency feature extractor. Its

trigger modules with 1×1 convolutions are excellent at capturing fine-grained textures and edges crucial for identifying irregular borders and granularity in melanomas. The squeeze phase utilizes 1×1 filters, while the expansion phase utilizes 1×1 and 3×3 filters. The squeeze operation's output layer $f\{y\}$ using the kernel w can be written as:

$$f\{y\} = \sum_{fm1=1}^{FM} \sum_{c=1}^{C} w_c^f x_c^{fm1} \quad (17)$$

The role of the DenseNet121 branch is feature reuse and propagation. Dense connections ensure that gradients flow directly to earlier layers, preserving intricate feature hierarchies that might be lost in very deep networks, thus improving the learning of complex lesion patterns. The DenseNet framework consists of the bottleneck layer and the dense block. A transition layer is added after each dense block to handle the downsampling process.

$$I_N = H_N[I_0, I_1, I_2, \dots, I_{N-1}] \quad (18)$$

where the feature maps from the preceding $N-1$ layers are represented by $[I_0, I_1, I_2, \dots, I_{N-1}]$ and are joined to the N^{th} layer, denoted by I_N.

The ResNet50 branch enables very deep representation learning via residual blocks. This branch is designed to learn more abstract, high-level semantic features about the overall structure and context of the lesion.

```
Algorithm 1: Proposed Algorithm
Input: Raw skin lesion image dataset
Output: Classification result (Benign /
  Malignant)
Step 1: Data Augmentation with GAN
  Sample a mini-batch of real images
  Sample a mini-batch of random noise
  vectors
  Generate synthetic images
  Update Discriminator D
  Update Generator G
  Augment dataset D with high-quality
  generated images
Step 2: Preprocessing
  For each image in the augmented dataset
  D do
    Apply Gaussian Blur
    Normalize using adaptive BN
  end for

Step 3: ROI segmentation with Deep
        Dilated-Focus U-Net
  For each preprocessed image do
    Encoder Path:
      Apply convolutional layers with ReLU
      & downsampling (MaxPool)
      Apply CBAM module after each
      convolution
    Bottleneck:
      Apply dilated convolutions to expand
      The receptive field
    Decoder Path:
      Apply transposed convolutions for
      upsampling
      Concatenate features from the
      encoder skip connections
      Apply the CBAM module for feature
      refinement
      Generate a binary segmentation mask
      via sigmoid activation
    Extract ROI
  end for
Step 4: Classification via TriBlindNet
  For each ROI image I_Ruz do
    Feature Extraction:
      Extract features using SqueezeNet
      Trigger modules
      Extract features using DenseNet121
      dense blocks
    Feature Fusion:
      Fuse feature maps from
      both branches via concatenation
    Classification:
      Pass fused features through fully
      connected layers
      Obtain class probabilities via
      softmax layer
      Assign class C = arg max(P)
  end for
```

III. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed method was benchmarked against established architectures. The simulation was conducted on NVIDIA CUDA-compatible GPUs to handle the computational load of the models. The framework was built in Python using TensorFlow, PyTorch, and Keras. Key aspects of the implementation included automated data preprocessing and augmentation, and the novel GazFOX optimization model for hyperparameter tuning. The model was trained and evaluated using a substantial dataset of 25,331 dermoscopic images. For the binary classification task, these images were grouped into a malignant class (4,852 images), comprising melanoma, basal cell carcinoma, actinic keratosis, and squamous cell carcinoma, and a benign class (20,479 images), consisting of melanocytic nevus, benign keratosis, dermatofibroma, and vascular lesions. A rigorous validation strategy was employed to ensure a robust and unbiased performance evaluation.

The entire dataset was first partitioned into a held-out test set (30% of the data), which was completely unseen during all stages of model development and used solely for the final reported results. The remaining 70% of the data was used for model training, using stratified 5-fold cross-validation to optimize hyperparameters and prevent overfitting, ensuring the model's generalizability before the final assessment on the held-out test set.

Key metrics for this analysis include:

- Number of Parameters: indicates the model's size and complexity, directly relating to its storage requirements and memory footprint.

- Floating Point Operations (FLOPs): measures the computational intensity required for a single forward pass, serving as a proxy for processing speed and power consumption.

Table I compares the computational cost of various models.

TABLE I. COMPUTATIONAL COMPLEXITY AND MODEL SIZE COMPARISON OF DIFFERENT ARCHITECTURES

Model	Number of Parameters	FLOPS
SqueezeNet	1.2 million parameters	1.7 billion FLOPS
DenseNet121	8 million parameters	2.8 billion FLOPS
ResNet50	23 million parameters	4.1 billion FLOPS
VGG16	23 million parameters	15.3 billion FLOPS
TriBlendNet	15 million parameters	3.4 billion FLOPS

SqueezeNet is the most efficient model with the fewest parameters and the lowest computational cost (FLOPS). VGG16 is the least efficient, requiring the most computational resources despite having the same number of parameters as ResNet50. DenseNet121 and TriBlendNet offer a middle ground, balancing model size and computational demand. The results in Table I demonstrate the notable efficiency of the proposed TriBlendNet architecture, showing that it achieves a favorable balance, incorporating the representational power of deeper networks while maintaining a computational footprint closer to that of more efficient models like DenseNet121.

A. Overall Performance Analysis

The proposed TriBlendNet model demonstrates superior performance on the training set across all metrics, achieving an accuracy of 98.59% and an F1-score of 98.4%. The high performance on training data suggests that TriBlendNet possesses strong representational power and is able to learn the relevant features from the training data effectively. Tables III and IV elucidate the comparative analysis of performance metrics across various models on a test dataset (20%) and a validation dataset (10% of the total data).

TABLE II. PERFORMANCE ON TRAINING DATASET (70%)

Model	Accuracy	Precision	Recall	F1-score	Specificity
SqueezeNet	0.89	0.88	0.89	0.908	0.895
DenseNet121	0.88	0.89	0.89	0.909	0.895
ResNet50	0.87	0.89	0.90	0.872	0.892
VGGNet	0.90	0.89	0.90	0.882	0.902
Proposed TriBlendNet	0.98	0.98	0.98	0.984	0.988

To further validate the design of TriBlendNet, an ablation study was conducted to evaluate the contribution of each branch individually and in combination. As shown in Table V, the SqueezeNet branch alone achieved an accuracy of 90.1%, highlighting its strength in capturing fine-grained textures and edges due to its efficient trigger modules. The DenseNet121 branch, with its dense connections, achieved 91.5% accuracy, demonstrating robust feature reuse and hierarchical learning. The ResNet50 branch, leveraging deep residual learning, achieved the highest individual performance (92.2%) among the three, indicating its capacity for learning high-level semantic features. Consistent improvements are observed when combining branches. The pairing of SqueezeNet and DenseNet121 reached 94.5% accuracy, while SqueezeNet with ResNet50 achieved 94.8%. The combination of DenseNet121 and ResNet50 yielded 95.6%, underscoring the complementary nature of hierarchical and semantic features. The full TriBlendNet model, integrating all three branches, achieved the best performance, confirming that the synergistic fusion of texture, hierarchy, and semantics is critical for superior skin cancer classification.

TABLE III. COMPARATIVE PERFORMANCE EVALUATION ON THE HELD-OUT TEST SET (20%)

Model	Accuracy	Precision	Recall	F-score	Specificity	Sensitivity
SqueezeNet	0.88	0.87	0.88	0.88	0.87	0.88
DenseNet121	0.87	0.86	0.88	0.87	0.87	0.89
ResNet50	0.86	0.88	0.87	0.86	0.87	0.87
VGG Net	0.87	0.88	0.89	0.87	0.88	0.87
Proposed TriBlendNet	0.97	0.97	0.97	0.97	0.97	0.97

TABLE IV. MODEL PERFORMANCE ON THE VALIDATION SUBSET (10%)

Model	Accuracy	Precision	Recall	F-score	Specificity	Sensitivity
SqueezeNet	0.86	0.842	0.868	0.86	0.85	0.87
DenseNet121	0.85	0.857	0.849	0.83	0.85	0.85
ResNet50	0.85	0.870	0.855	0.84	0.84	0.85
VGGNet	0.84	0.852	0.869	0.85	0.87	0.86
Proposed TriBlendNet	0.985	0.964	0.963	0.96	0.96	0.96

TABLE V. ABLATION STUDY ON THE IMPACT OF PIPELINE COMPONENTS ON CLASSIFICATION

Model configuration	Accuracy	Precision	Recall	F1-score
SqueezeNet branch only	0.90	0.89	0.90	0.89
DenseNet121 branch only	0.91	0.90	0.91	0.91
ResNet50 branch only	0.92	0.91	0.92	0.92
SqueezeNet+DenseNet121	0.94	0.94	0.94	0.94
SqueezeNet+ResNet50	0.94	0.94	0.95	0.94
DenseNet121+ResNet50	0.95	0.95	0.95	0.95

The TriBlendNet architecture leverages three parallel branches:

- SqueezeNet branch: Focuses on high-frequency texture features using 1×1 convolutions.
- DenseNet121 branch: Employs dense connections for hierarchical feature propagation.
- ResNet50 branch: Utilizes residual learning for high-level semantic representation.

Features from all branches are concatenated and passed through a fully connected layer for final classification.

The results in Table VI demonstrate that the proposed model consistently outperforms existing architectures.

TABLE VI. PERFORMANCE COMPARISON OF THE WITH STATE-OF-THE-ART ARCHITECTURES

Model	Accuracy	Precision	Recall	F1-score
EfficientNet-V2 (Small) [12]	0.95	0.94	0.95	0.95
ConvNeXt (Tiny) [13]	0.95	0.96	0.95	0.95
ViT-Base/16 [14]	0.94	0.94	0.94	0.94
Proposed TriBlendNet	0.97	0.97	0.97	0.97

The proposed TriBlendNet was also tested on the HAM10000 dataset [10], which consists of 10,015 dermatoscopic images curated into 7 diagnostic classes validated by histopathologists. Table VII shows the results of the proposed and other models on this dataset, demonstrating that it consistently outperforms existing architectures.

TABLE VII. MULTI-CLASS CLASSIFICATION PERFORMANCE ON THE HAM10000 DATASET TEST SET

Model	Accuracy	Precision	Recall	F1-Score
SqueezeNet	0.82	0.78	0.75	0.76
DenseNet121	0.84	0.81	0.80	0.80
ResNet50	0.83	0.80	0.79	0.79
VGG16	0.83	0.79	0.78	0.78
EfficientNet-V2 (S)	0.86	0.83	0.83	0.83
Proposed TriBlendNet	0.92	0.90	0.89	0.89

This work introduces significant innovations through a novel and synergistic three-branch CNN architecture, TriBlendNet, which is designed to harness the complementary strengths of distinct CNN families rather than being a simple ensemble, a fact proven by the ablation study. This is complemented by the development of the Deep-dilated Focus U-Net, a segmentation model that uniquely integrates dilated convolutions with attention mechanisms (CBAM) to achieve superior ROI extraction. These components form a holistically optimized, end-to-end pipeline that incorporates advanced techniques, such as GAN-augmentation and adaptive BN, which together enable the proposed framework to achieve state-of-the-art results.

IV. CONCLUSION

Given that skin cancer is both highly prevalent and potentially lethal, developing advanced detection methods is critical to improving patient outcomes. This study presents TriBlendNet, an innovative framework that combines advanced architectures and techniques for early detection of skin cancer, representing a significant step forward. The performance evaluation of TriBlendNet demonstrates an average accuracy of 98.59%, surpassing traditional models such as DenseNet121, SqueezeNet, ResNet50, and VGGNet. These results highlight its strong capability to correctly identify both positive and negative cases while preserving high sensitivity and specificity, thereby reducing diagnostic errors. Overall, TriBlendNet achieves substantial improvements over existing methods across multiple performance metrics.

Future directions for TriBlendNet include expanding dataset diversity, integrating explainable AI techniques, optimizing computational efficiency and scalability, and exploring federated learning approaches to enable collaborative advances in automated skin cancer detection.

REFERENCES

[1] N. Zhang, Y. X. Cai, Y. Y. Wang, Y. T. Tian, X. L. Wang, and B. Badami, "Skin cancer diagnosis based on optimized convolutional neural network," *Artificial Intelligence in Medicine*, vol. 102, Jan. 2020, Art. no. 101756, https://doi.org/10.1016/j.artmed.2019.101756.

[2] S. K. Singh, V. Abolghasemi, and M. H. Anisi, "Fuzzy Logic with Deep Learning for Detection of Skin Cancer," *Applied Sciences*, vol. 13, no. 15, Aug. 2023, Art. no. 8927, https://doi.org/10.3390/app13158927.

[3] S. Sikandar, R. Mahum, A. E. Ragab, S. Y. Yayilgan, and S. Shaikh, "SCDet: A Robust Approach for the Detection of Skin Lesions," *Diagnostics*, vol. 13, no. 11, May 2023, Art. no. 1824, https://doi.org/10.3390/diagnostics13111824.

[4] M. Tahir, A. Naeem, H. Malik, J. Tanveer, R. A. Naqvi, and S. W. Lee, "DSCC_Net: Multi-Classification Deep Learning Models for Diagnosing of Skin Cancer Using Dermoscopic Images," *Cancers*, vol. 15, no. 7, Apr. 2023, Art. no. 2179, https://doi.org/10.3390/cancers15072179.

[5] J. Qadir, "Enhancing Skin Disease Diagnosis: A Hybrid Approach Combining Vision Transformer and Feature Selection Techniques.," *Zanin Journal of Science and Engineering*, vol. 1, no. 1, pp. 54–71, Mar. 2025, https://doi.org/10.64362/zjse.37.

[6] S. B. Mukadam and H. Y. Patil, "Skin Cancer Classification Framework Using Enhanced Super Resolution Generative Adversarial Network and Custom Convolutional Neural Network," *Applied Sciences*, vol. 13, no. 2, Jan. 2023, Art. no. 1210, https://doi.org/10.3390/app13021210.

[7] A. Hekler *et al.*, "Using Multiple Dermoscopic Photographs of One Lesion Improves Melanoma Classification via Deep Learning: A Prognostic Diagnostic Accuracy Study," *Journal of the American Academy of Dermatology*, vol. 90, no. 5, pp. 1028–1031, May 2024, https://doi.org/10.1016/j.jaad.2023.11.065.

[8] A. E. Mrabet, M. Benaly, I. Alihamidi, B. Kouach, L. Hlou, and R. E. Gouri, "Enhancing Early Detection of Skin Cancer in Clinical Practice with Hybrid Deep Learning Models," *Engineering, Technology & Applied Science Research*, vol. 15, no. 2, pp. 20927–20933, Apr. 2025, https://doi.org/10.48084/etasr.9753.

[9] International Skin Imaging Collaboration, "SIIM-ISIC 2020 Challenge Dataset." International Skin Imaging Collaboration, 2020, https://doi.org/10.34970/2020-DS01.

[10] "Skin Cancer MNIST: HAM10000." Kaggle, [Online]. Available: https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000.

[11] A. T. Priyeshkumar, G. Shyamala, T. Vasanth, and P. V. Selvan, "Transforming Skin Cancer Diagnosis: A Deep Learning Approach with the Ham10000 Dataset," *Cancer Investigation*, vol. 42, no. 10, pp. 801–814, Nov. 2024, https://doi.org/10.1080/07357907.2024.2422602.

[12] J. Amin, M. Azhar, H. Arshad, A. Zafar, and S. H. Kim, "Skin-lesion segmentation using boundary-aware segmentation network and classification based on a mixture of convolutional and transformer neural networks," *Frontiers in Medicine*, vol. 12, Mar. 2025, Art. no. 1524146, https://doi.org/10.3389/fmed.2025.1524146.

[13] M. Tan and Q. V. Le, "EfficientNetV2: Smaller Models and Faster Training," 2021, https://doi.org/10.48550/ARXIV.2104.00298.

[14] Z. Liu, H. Mao, C. Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s." arXiv, 2022, https://doi.org/10.48550/ARXIV.2201.03545.

[15] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." arXiv, 2020, https://doi.org/10.48550/ARXIV.2010.11929.

[16] I. J. Goodfellow *et al.*, "Generative Adversarial Networks." arXiv, 2014, https://doi.org/10.48550/ARXIV.1406.2661.

[17] M. M. Musthafa, T. R. Manesh, V. V. Kumar, and S. Guluwadi, "Enhanced skin cancer diagnosis using optimized CNN architecture and checkpoints for automated dermatological lesion classification," *BMC*

Medical Imaging, vol. 24, no. 1, Aug. 2024, Art. no. 201, https://doi.org/10.1186/s12880-024-01356-8.

[18] 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*, vol. 2024, pp. 1–18, Feb. 2024, https://doi.org/10.1155/2024/3022192.

[19] M. F. Aslan, "Comparison of vision transformers and convolutional neural networks for skin disease classification," in *Proceedings of the International Conference on New Trends in Applied Sciences*, 2023, vol. 1, pp. 31–39.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation." arXiv, 2015, https://doi.org/10.48550/ARXIV.1505.04597.

[21] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module." arXiv, July 18, 2018, https://doi.org/10.48550/arXiv.1807.06521.