Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 3, 2026

IMPROVED MELANOMA DETECTION USING IMAGE SEGMENTATION AND A LIGHTWEIGHT CNN FRAMEWORK

Iqra Khan[*1], Muhammad Javed Iqbal[2]

[*1,2]*Department of Computer Science, Faculty of Telecommunication and Information Engineering University of Engineering and Technology, Taxila, Pakistan*

[*1]ik9638047@gmail.com, [2]javed.iqbal@uettaxila.edu.pk

DOI: https://doi.org/10.5281/zenodo.21485124

Keywords
Melanoma Detection, Dermoscopic Image Segmentation, Skin Cancer, Convolutional Neural Network, Hybrid Segmentation

Article History
Received: 01 February 2026
Accepted: 17 March 2026
Published: 31 March 2026

Corresponding Author: *
Iqra Khan

Abstract
The most severe type of skin cancer is melanoma, and early diagnosis is so important for patients' survival. Manual dermoscopic examination is time-consuming, subjective and impractical for large-scale screening, and existing computational methods have significant drawbacks: handcrafted feature extraction methods do not transfer well between datasets, and deep learning methods like ResNet-50 and Inception-ResNet-v2 achieve higher accuracy but require significant computational resources, unsuitable in the clinic. Furthermore, segmentation and classification are generally considered as two independent processes, and class imbalance in dermoscopic data sets is seldom addressed, and may decrease the accuracy of melanoma detection. In this paper, a two-stage approach deep learning-based automated melanoma detection system is proposed for ISIC 2018 Lesion Segemtation and Classifiation dataset with minimum computational complexity. Preprocessing techniques used were the removal of hair artifacts, the application of a CLAHE contrast enhancement, and pixel normalization; the class imbalance was handled by class augmentation and loss weight. A Mobile-UNet architecture was used to segment lesions followed by a two-stage trained EfficientNet-B0 classification network. Its Dice Coefficient of 87.8%, IoU of 80.6%, classification accuracy of 92.5%, F1-score of 91.8% and AUC-ROC of 95.1% outperform six benchmark methods such as SVM, Random Forest, VGG-16, ResNet-50, Inception-V3, and MobileNetV2-UNet, indicating its feasibility for efficient, real-world clinical utilization.

I. INTRODUCTION
Skin cancer is among the fastest-growing cancers worldwide, and although melanoma accounts for only a small fraction of skin cancer cases, it is responsible for the majority of skin-cancer-related deaths [1]. Early diagnosis raises the five-year survival rate above 95%, whereas late diagnosis reduces it to below 20%, underscoring the clinical importance of timely detection.

Dermoscopy remains the clinical gold standard for melanoma diagnosis, using magnified polarized light to reveal subsurface skin structures invisible to the naked eye. Dermatologists typically apply the ABCDE criteria Asymmetry, Border irregularity, Color variation, Diameter, and Evolution to flag suspicious lesions. Manual interpretation, however, is time-consuming, requires expertise that is scarce in rural or under-resourced settings, and is prone to inter- and intra-observer variability that can delay diagnosis.

Computer-aided diagnosis (CAD) based on Convolutional Neural Networks (CNNs) has therefore been widely explored as an alternative to manual visual inspection. Basic CNN classifiers have reported accuracies above 90% on dedicated melanoma databases [3], and large public repositories such as HAM10000 have enabled increasingly sophisticated models [4]. Handcrafted-feature pipelines using descriptors such as the Histogram of Oriented Gradients (HOG) with Support Vector Machines (SVM) have, in some cases, matched or exceeded early CNN-based methods [5], while other approaches combine patient metadata with gradient-boosted classifiers such as XGBoost and synthetic oversampling to counter class imbalance [6].

Despite this progress, three limitations recur in the literature. First, several high-accuracy systems rely on computationally expensive components — Capsule Networks with metaheuristic tuning [7], Inception-ResNet-v2 combined with Vision Transformers [8], or HRNet features passed through a Deep Belief Network [9] — that are difficult to deploy on mobile or point-of-care hardware. Second, lesion segmentation and classification are frequently treated as separate problems, although two-stage detection-then-classification pipelines using networks such as YOLOv3 or Mask R-CNN have shown promise [10]. Third, the severe class imbalance typical of dermoscopic datasets is addressed inconsistently — through SMOTE [6], GAN-based augmentation [14], or class weighting — with little attention paid to combining these techniques with genuinely lightweight architectures suitable for resource-limited clinical settings.

To address these gaps, this paper proposes a compact, two-stage deep learning system that integrates lesion segmentation with lesion classification. Dermoscopic images from ISIC 2018 are first preprocessed using black-hat morphological hair removal with Telea inpainting, followed by Contrast Limited Adaptive Histogram Equalization (CLAHE) and pixel normalization. A lightweight Mobile-UNet model, built on a MobileNetV2 encoder, then segments the lesion region. In the second stage, the segmented lesion is classified as benign or malignant by a fine-tuned EfficientNet-B0 backbone, trained with a two-phase strategy in which the classification head is trained first with the backbone frozen, followed by end-to-end fine-tuning. Class-weighted loss and data augmentation are applied throughout to mitigate class imbalance. The proposed framework is evaluated using Accuracy, Precision, Recall, F1-Score, Specificity, and AUC-ROC, and is compared against six methods reported in the literature

.

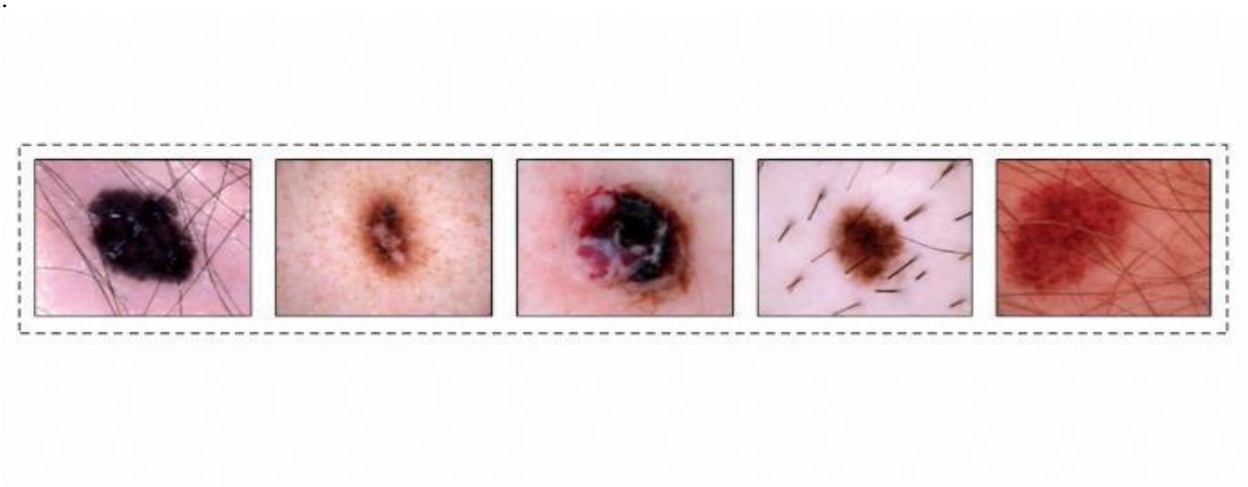


Fig. 1. Sample dermoscopic images of melanoma skin lesions from the HAM10000 dataset.

This work makes the following contributions:

1) A two-stage, computationally efficient

melanoma detection system combining Mobile-UNet segmentation with EfficientNet-B0 classification.
2) A dermoscopic-image preprocessing pipeline integrating hair removal, CLAHE contrast enhancement, and pixel normalization.
3) A class-imbalance mitigation strategy combining data augmentation with a class-weighted loss function.
4) An experimental comparison against six literature methods, demonstrating improved accuracy and AUC-ROC while remaining suitable for resource-constrained clinical deployment.
The remainder of this paper is organized as follows. Section II reviews related work on handcrafted, deep-learning, and hybrid segmentation–classification approaches for melanoma detection. Section III describes the proposed methodology. Section IV presents the experimental results and comparative analysis. Section V concludes the paper and outlines future work.

II. RELATED WORK

Prior work on automated melanoma detection can be grouped into handcrafted-feature methods, CNN-based classifiers, attention/transformer models, segmentation-focused methods, and two-stage segmentation–classification pipelines, the last of which is closest to the present work.
Among handcrafted and classical approaches, PCA-based feature reduction with SMOTE and SVM reached 99.15% accuracy on HAM10000, outperforming baseline ResNet-50 and MobileNet models [5]. Combining CNN features with patient metadata, SMOTE balancing, and an XGBoost classifier achieved 94.6% accuracy and an AUC of 0.970 on a hybrid ISIC dataset, though specificity dropped for larger lesions [6]. CNN-extracted features from GoogLeNet and ResNet-50, paired with SVM, Decision Tree, and KNN classifiers, reached 98.44% accuracy on ISBI-2017 [15].
Pure CNN classifiers have also performed strongly: a five-layer CNN trained on a proprietary dataset achieved 91% accuracy [3], while depth studies found that four convolutional layers outperformed five before overfitting occurred [16]. A Fully Convolutional Deep Skin-Cancer Network (FCDS-CNN) outperformed pretrained ResNet152V2, EfficientNetV2B0, MobileNetV3, and InceptionResNetV2 baselines with 96.66% accuracy on HAM10000 [4]. Attention- and transformer-based models such as ASCU-Net (92.6% accuracy) and a ConvNeXt variant with spatial attention (94.0% accuracy) improved boundary sensitivity at higher computational cost, while a Vision Transformer combined with U-Net segmentation reached 98.65% accuracy on ISIC-2020 [8].
For segmentation-focused approaches, test-time augmentation applied to UNet/ResUNet variants achieved an 85.44% Jaccard Index, exceeding the 78.6% dermatologist inter-observer agreement reported for the same task [22], and multi-detector ensembles (YOLOv3, Faster R-CNN, RetinaNet) combined with U-Net segmentation achieved a mean Average Precision of 0.835 on mobile-phone images, at the cost of higher runtime [23].
The two-stage pipelines most closely related to this work are summarized in Table I. They include a YOLOv3 detector with SegNet segmentation that did not account for hair or bubble artifacts [10]; an ADL-MDC framework combining CLAHE, K-Means segmentation, a Capsule Network, and a Crow-Search-optimized Sparse Autoencoder, bottlenecked by its clustering-based segmentation stage [7]; a hybrid FrCN–U-Net with an R-CNN classifier operating at high computational cost [11]; a Mask R-CNN and ResNet50 pipeline affected by a train–test class-imbalance mismatch [12]; and a clustering-guided EfficientNet-B6 framework evaluated on a large 48,007-image dataset [25].

TABLE I
Summary of Closely Related Two-Stage Segmentation–Classification Studies

Ref.	Segmentation / Classification Approach	Dataset	Reported Performance
[10]	Fine-tuned YOLOv3 detector + fine-tuned SegNet segmentation	ISIC	96% mAP (detection); 95.16% accuracy (segmentation)
[7]	CLAHE + K-Means segmentation, Capsule Network + Sparse Autoencoder (Crow Search optimized) classifier	Dermoscopic (ADL-MDC)	98.27% accuracy
[11]	Hybrid FrCN–U-Net segmentation + R-CNN classifier	ISBI-2017 / ISIC-2018	95.8% seg. accuracy; 94.44% classification accuracy
[12]	Mask R-CNN segmentation + ResNet50 classifier	ISIC-2020	95.49% seg. accuracy; 96.75% classification accuracy
[9]	Ensemble segmentation + HRNet-based Deep Belief Network classifier (Skin-DeepNet)	ISIC-2019 / HAM10000	Up to 99.93% (ISIC-2019); 99.65–100% (HAM10000)
[25]	Clustering-guided EfficientNet-B6 (k-means, t-SNE, hierarchical clustering, NCA)	ISIC-2019/2020 (48,007 images)	99.62% AUC

Overall, a systematic review of melanoma-detection networks covering 2015–2021 found that CNN and hybrid models generally outperform dermatologist visual assessment (around 84% accuracy), but that the most accurate systems typically rely on components ill-suited to point-of-care deployment, and that class-imbalance techniques are rarely combined with genuinely lightweight architectures [19]. These observations motivate the present work, which builds on a MobileNetV2-based U-Net for segmentation and a transfer-learning EfficientNet-B0 classifier with class-weighted loss to obtain a lightweight, deployable pipeline.

III. RESEARCH METHODOLOGY

A. Overview of the Proposed System

The proposed melanoma detection system adopts a two-stage deep learning pipeline that separates lesion localization from disease classification, allowing each stage to be optimized independently while keeping the overall pipeline lightweight. Images are first acquired from the ISIC 2018 dataset and passed through a preprocessing and class-balancing phase. Stage 1 uses a MobileNetV2-UNet model to segment the lesion area pixel-wise, separating it from irrelevant background and artifacts. The segmented lesion image is then passed to Stage 2, a fine-tuned EfficientNetB0 model, which estimates the probability of malignancy and outputs a binary decision: benign or melanoma. The overall pipeline is shown in Fig. 2.

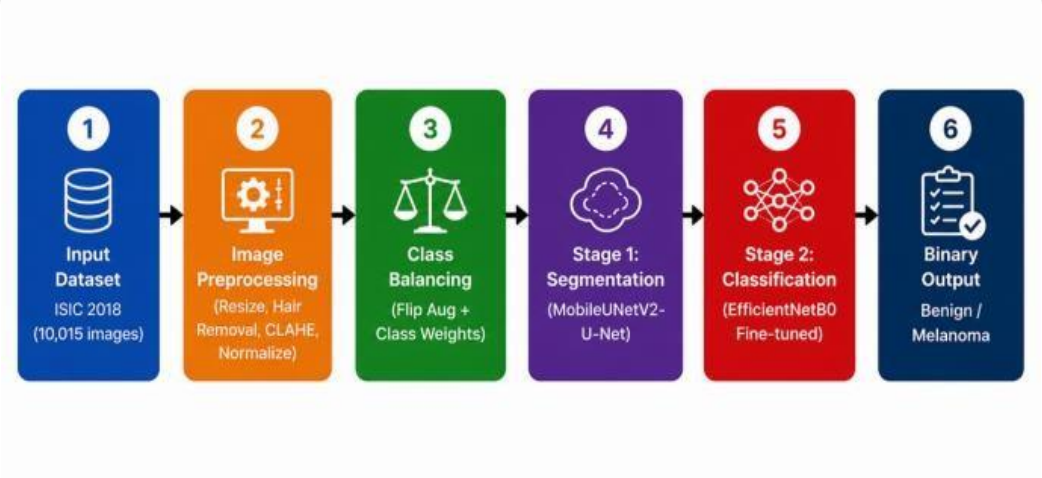


Fig. 2. Proposed two-stage melanoma detection pipeline.

B. Dataset

The dataset used in this study is a publicly accessible benchmark provided by the International Skin Imaging Collaboration (ISIC) for skin lesion classification (Task 3). The full dataset contains seven diagnostic categories; this work addresses a binary classification problem using two classes — Melanocytic Nevi (benign) and Melanoma (malignant). A subset of the images was preprocessed, scored, and balanced to obtain equal numbers of benign and melanoma samples. An 80:20 stratified split was used to create training and validation sets, with a separate holdout set reserved for final testing. Table II summarizes the dataset split used in this work.

TABLE II
ISIC 2018 Dataset Split

Split	Total	Benign	Melanoma	Melanoma %
Training	420	210	210	50.0%
Validation	60	30	30	50.0%
Test (Holdout)	120	60	60	50.0%
Total	~600	300	300	50.0%

C. Image Preprocessing

Before segmentation and classification, all dermoscopic images are preprocessed to standardize image quality and remove clinically irrelevant artifacts, as illustrated in Fig. 3. Images are resized to 224×224 pixels using bilinear interpolation, matching the default ImageNet input size used by the MobileNetV2 and EfficientNetB0 backbones. Hair artifacts, a common source of boundary noise in dermoscopic images, are removed by applying a BlackHat morphological filter with a 3×3 elliptical structuring element and a threshold value of 10; pixels above this threshold are treated as hair and inpainted using the Telea algorithm with an inpainting radius of 3 pixels. Contrast is then enhanced in the LAB color space using Contrast Limited Adaptive Histogram Equalization (CLAHE), which improves lesion-boundary visibility while limiting noise amplification. Finally, pixel intensities are rescaled to the [0, 1] range, matching the range expected by the pretrained ImageNet weights and accelerating gradient-based optimization.

Fig. 3. Image preprocessing pipeline applied to dermoscopic images.

D. Class Balancing

The ISIC 2018 subset used in this work is imbalanced, with melanoma images forming the minority class. Classifiers trained directly on such data tend to be biased toward the majority (benign) class, often yielding high overall accuracy but low melanoma sensitivity. Two complementary strategies are used to address this. First, melanoma training images are horizontally flipped, effectively doubling the number of melanoma samples during training. Second, inverse-frequency class weights are computed using scikit-learn's compute_class_weight utility and passed to the loss function, so that misclassifications of the minority melanoma class are penalized more strongly in a class-weighted binary cross-entropy loss.

E. Stage 1: Lesion Segmentation (MobileNetV2-UNet)

The first stage performs pixel-wise lesion segmentation, producing a binary mask that separates the skin from the lesion. Ground-truth training masks are generated from the preprocessed grayscale image using Otsu's global thresholding, followed by morphological closing and opening with a 3×3 square structuring element.

The segmentation model is a U-Net with a MobileNetV2 encoder pretrained on ImageNet, as shown in Fig. 4. Multi-scale feature maps are extracted at four spatial resolutions from the encoder — block_1_expand_relu (112×112×96), block_3_expand_relu (56×56×144), block_6_expand_relu (28×28×192), and block_13_expand_relu (14×14×576) — which form the bottleneck bridge. The decoder comprises four transposed-convolution blocks of 256, 128, 64, and 32 filters, respectively; each block upsamples its feature map by a factor of 2, concatenates it with the corresponding encoder skip connection, and passes the result through a Conv2D layer with ReLU activation and batch normalization. A final 1×1 convolution with sigmoid activation produces the binary segmentation mask. The model is optimized with Adam (learning rate 1×10^{-4}) using binary cross-entropy loss, a batch size of 16, and a maximum of 20 epochs, with EarlyStopping (patience = 10) and ReduceLROnPlateau (factor = 0.5, patience = 5) callbacks. Segmentation quality is assessed using the Dice coefficient and Intersection-over-Union (IoU), defined in (1) and (2):

$$Dice = \frac{2|Y \cap \hat{Y}| + \varepsilon}{|Y| + |\hat{Y}| + \varepsilon} \quad (1)$$

$$IoU = \frac{|Y \cap \hat{Y}| + \varepsilon}{|Y \cup \hat{Y}| + \varepsilon} \quad (2)$$

where Y is the ground-truth mask, $\hat{Y}$ is the predicted binary mask thresholded at 0.5, and $\varepsilon = 1\times10^{-7}$ is a smoothing constant that prevents division by zero.

the melanoma probability $p \in [0,1]$. The classification decision is made according to (3):

$$Class(p) = \begin{cases} Melanoma, & p > 0.5 \\ Benign, & otherwise \end{cases} \quad (3)$$

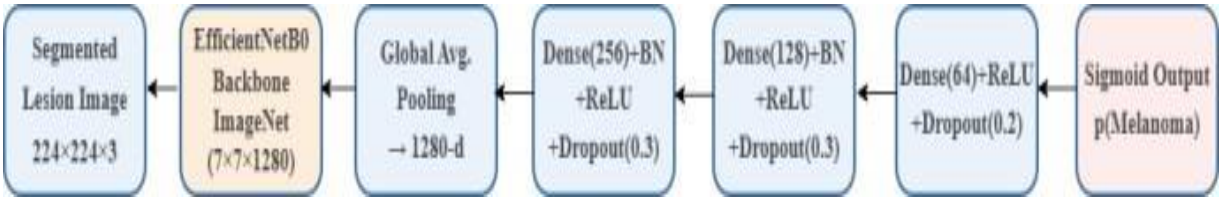


Fig. 5. EfficientNetB0 backbone with the proposed custom classification head.

Training follows a two-phase transfer-learning approach to avoid catastrophic forgetting of ImageNet-learned features. In Phase 1 (feature extraction), the EfficientNetB0 backbone is frozen and only the classification head is trained using the Adam optimizer (learning rate 1×10^{-3}), binary cross-entropy loss, and a batch size of 32, for up to 30 epochs. In Phase 2 (fine-tuning), all backbone layers are unfrozen and the whole network is fine-tuned end-to-end at a reduced learning rate of 1×10^{-5} for up to 15 additional epochs, allowing the network to adapt to dermoscopic features without destabilizing gradient updates. EarlyStopping (patience = 10, best-weight restoration) and ReduceLROnPlateau (factor = 0.5, patience = 5) callbacks are used in both phases, along with class weighting to balance any residual class imbalance.

G. Hyperparameter Configuration

Table III summarizes the complete hyperparameter configuration used for the segmentation and classification models.

TABLE III

Hyperparameter Configuration for Both Models

Hyperparameter	Segmentation (Mobile-UNet)	Classification (EfficientNetB0)
Input Size	224×224×1	224×224×3
Backbone	MobileNetV2 (ImageNet)	EfficientNetB0 (ImageNet)
Backbone Training	Trainable (Full)	Frozen → Unfrozen
Optimizer	Adam	Adam
Initial LR	1×10^{-4}	$1\times10^{-3} \rightarrow 1\times10^{-5}$
Loss Function	Binary Cross-Entropy	Binary Cross-Entropy
Batch Size	16	32
Max Epochs	20	30 → 15 (Fine-Tune)
Dropout Rate	0.3, 0.3, 0.2	0.3, 0.3, 0.2
Output Activation	Sigmoid	Sigmoid

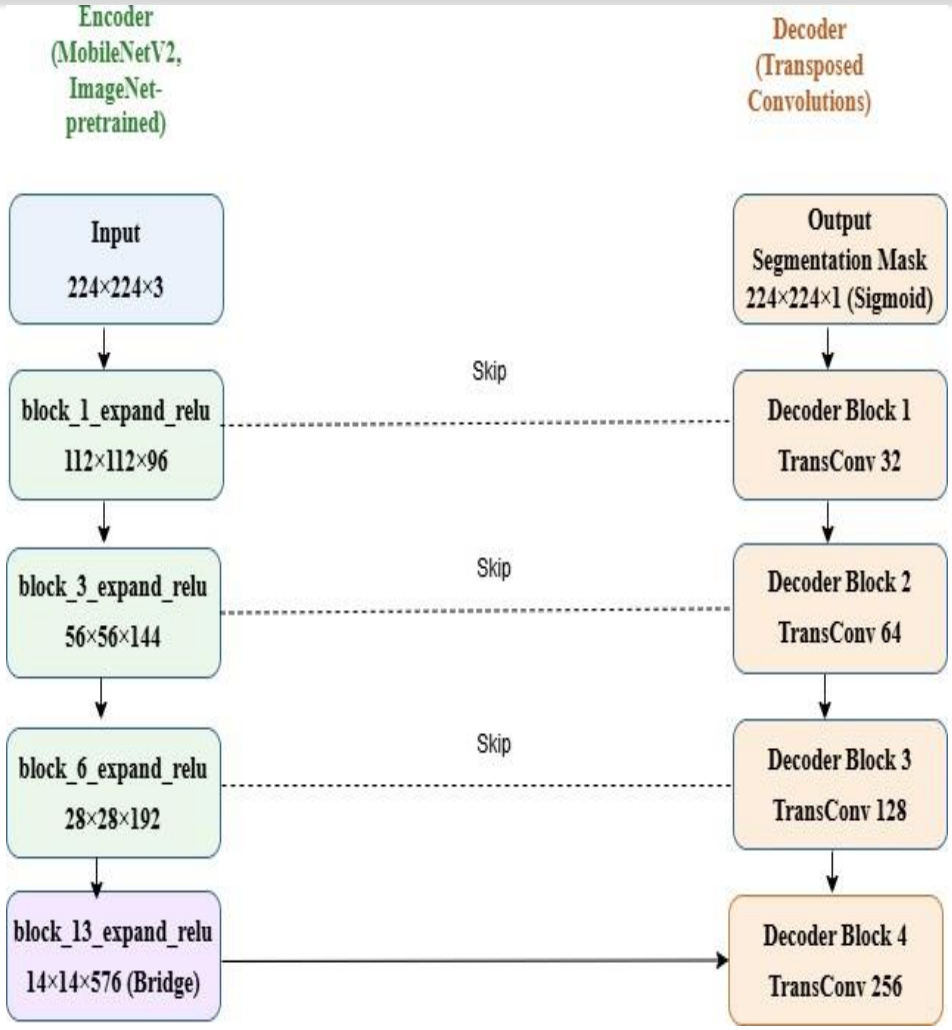


Fig. 4. MobileNetV2-UNet encoder–decoder architecture with skip connections for lesion segmentation

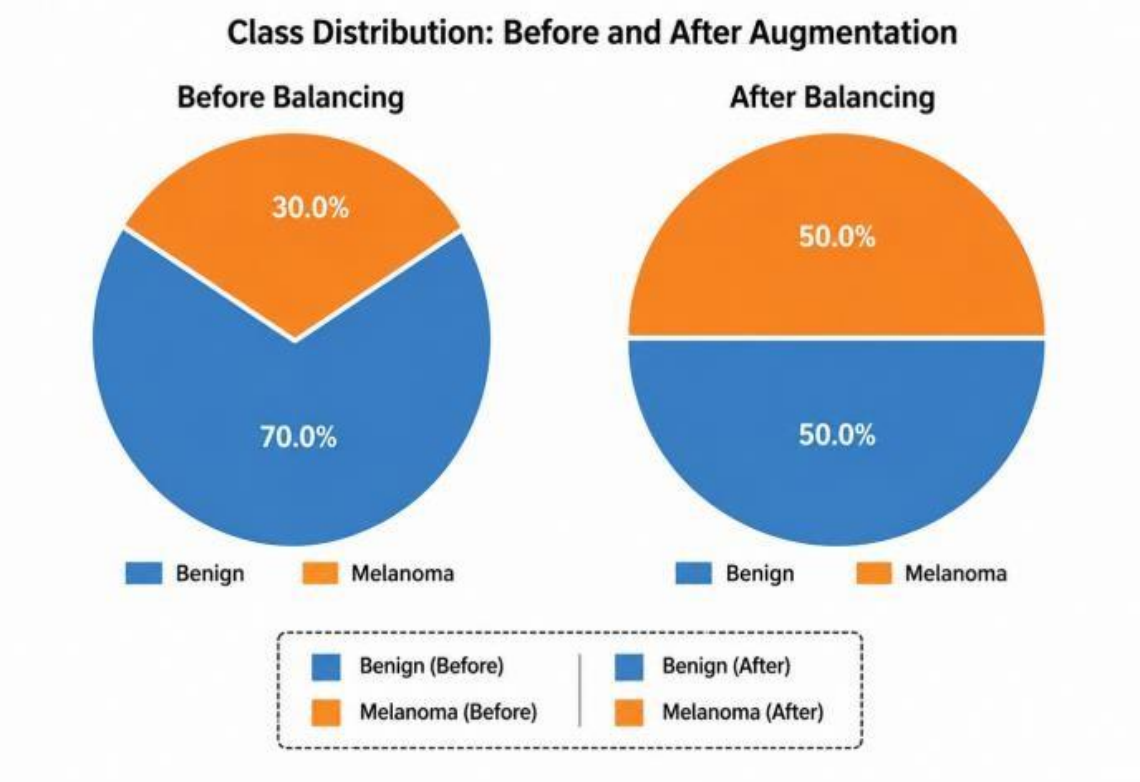
F. Stage 2: Melanoma Classification (EfficientNetB0)

The segmented lesion images produced by Stage 1 are fed into an EfficientNetB0-based binary classifier. Compared with earlier EfficientNet variants, EfficientNetB0 provides a favorable accuracy–efficiency trade-off by jointly scaling network depth, width, and input resolution, and by using mobile inverted bottleneck convolution (MBConv) blocks to convert the 224×224×3 input into a 7×7×1280 feature volume. A custom classification head is appended to the ImageNet-pretrained backbone, consisting of a Global Average Pooling layer that reduces the feature volume to a 1280-dimensional vector, followed by three fully connected blocks — Dense(256)–BatchNorm–ReLU–Dropout(0.3), Dense(128)–BatchNorm–ReLU–Dropout(0.3), Dense(64)— and a final Dense layer with sigmoid activation producing

H. Step-by-Step Algorithm

Algorithm 1: Two-Stage Melanoma Detection and Classification Pipeline
Input: Dermoscopic image set D from the ISIC 2018 directory
Output: Trained models M_seg, M_cls; evaluation metrics
1: D ← LoadImages(ISIC_2018_directory)

Algorithm 1 summarizes the complete two-stage melanoma detection and classification procedure.



```
2: for each image I ∈ D do
3:   I ← Resize(I, 224×224)
4:   I ← RemoveHairArtifacts(I)       D BlackHat +
Telea inpainting
5:   I ← ApplyCLAHE(I)                D LAB color
space
6:   I ← Normalize(I)                 D scale to
[0, 1]
7: end for
8: for each image I ∈ D do
9:   mask(I) ←
MorphOpen(MorphClose(OtsuThreshold(I)))
10: end for
11: D_bal ← BalanceClasses(D)          D
horizontal-flip on melanoma class
12: w ← ComputeClassWeights(D_bal)     D scikit-
learn
13: M_seg ← BuildMobileUNet(encoder =
MobileNetV2[ImageNet])
14: Train M_seg on D_bal: Adam(lr=1e-4), BCE loss,
batch=16, epochs≤20,
     EarlyStopping, ReduceLROnPlateau
15: M_cls ← BuildEfficientNetB0Classifier(backbone
= EfficientNetB0[ImageNet])
16: Freeze(M_cls.backbone)             D Phase 1:
feature extraction
17: Train M_cls.head on D_bal: Adam(lr=1e-3),
batch=32, epochs≤30, weighted loss w
18: Unfreeze(M_cls.backbone)           D Phase 2:
fine-tuning
19: Train M_cls on D_bal: Adam(lr=1e-5), batch=32,
epochs≤15, weighted loss w
20: for each image I ∈ ValidationSet do
21:   p ← M_cls(M_seg(I))
22:   prediction(I) ← Melanoma if p > 0.5, else
Benign
23: end for
24: Compute Accuracy, Precision, Recall, F1,
Specificity, AUC-ROC, ConfusionMatrix
25: return M_seg, M_cls, EvaluationMetrics
```

Fig. 6. Class distribution before and after dataset balancing

Fig. 6 shows that the original dataset had an imbalance of nearly 70% benign and 30% melanoma samples. To correct this, the melanoma class was augmented through horizontal flipping, producing a perfectly balanced set of 200 melanoma and 200 benign images prior to training.

B. Evaluation Metrics

The performance of the proposed system is evaluated using the following standard metrics:
Accuracy — the proportion of correctly classified samples over the total number of samples:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Precision — the ratio of true positives to all positive predictions:

IV.EXPERIMENTAL DETAILS AND RESULTS

A. Environmental Setup

All experiments were conducted in a GPU-enabled Google Colaboratory environment supported by an Intel Core i7 processor (2.80 GHz) with 16 GB RAM. The implementation was written in Python 3.8 using the TensorFlow 2.x/Keras API, OpenCV, scikit-learn, NumPy, Pandas, Matplotlib, and Seaborn for preprocessing, analysis, and visualization. The proposed system was trained and tested on the ISIC 2018 Task 3 benchmark, downloaded from the ISIC Archive. A total of 500 images were selected, preprocessed, and balanced through augmentation to 400 images (200 malignant and 200 benign).

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

Recall (Sensitivity) – the percentage of melanoma cases correctly detected among all melanoma cases:

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

F1-Score – the harmonic mean of Precision and Recall, useful for imbalanced datasets:

$$F1 - Score = \frac{2\times(Precision\times Recall)}{Precision+Recall} \quad (7)$$

Specificity – the model's ability to correctly identify benign (non-melanoma) cases:

$$Specificity = \frac{TN}{TN+FP} \quad (8)$$

AUC-ROC – the area under the Receiver Operating Characteristic curve, assessing overall discriminative ability:

$$TPR = \frac{TP}{TP+FN} \quad (9)$$

$$FPR = \frac{FP}{FP+TN} \quad (10)$$

weighted loss and two-phase training (frozen backbone followed by full fine-tuning). Three independent runs were performed for stability; Table V reports the mean values across these runs.

C. Experimental Results

$$AUC = \int_0^1 TPRd(FPR) \quad (11)$$

TABLE V

Classification Results Across Three Experiments

Metric	Exp1 (%)	Exp2 (%)	Exp3 (%)	Avg (%)
Accuracy	91.0	93.0	93.5	92.5
Precision	90.5	91.8	91.7	91.3
Recall	91.0	93.0	93.0	92.3
F1-Score	90.7	92.4	92.3	91.8
Specificity	90.0	91.5	92.0	91.2
AUC-ROC	94.0	95.5	95.8	95.1

Preprocessing and Hair Removal: All dermoscopic images were preprocessed using the pipeline described in Section III before training. Images were resized to 224×224 pixels, and hair artifacts were removed using black-hat morphological transformation with Telea inpainting, followed by CLAHE contrast enhancement in the LAB color space and normalization of pixel values to [0,1]. Hair removal improved the quality of segmentation masks by eliminating artifacts that could otherwise be confused with lesion boundaries, and CLAHE improved boundary visibility while limiting noise amplification.

1) Lesion Segmentation Results (Mobile-UNet): The Mobile-UNet segmentation model, built on a MobileNetV2 encoder pretrained on ImageNet, was trained for 20 epochs with an initial learning rate of 1×10^{-4}, using binary cross-entropy loss and the Adam optimizer. The segmentation performance for two experimental runs is summarized in Table IV.

TABLE IV

Segmentation Model Performance Results

Metric	**Exp 1** (%)	**Exp 2** (%)	**Avg.** (%)
Dice Coefficient	87.5	88.1	87.8
IoU Coefficient	80.2	81.0	80.6
Segmentation Acc.	91.3	92.0	91.7

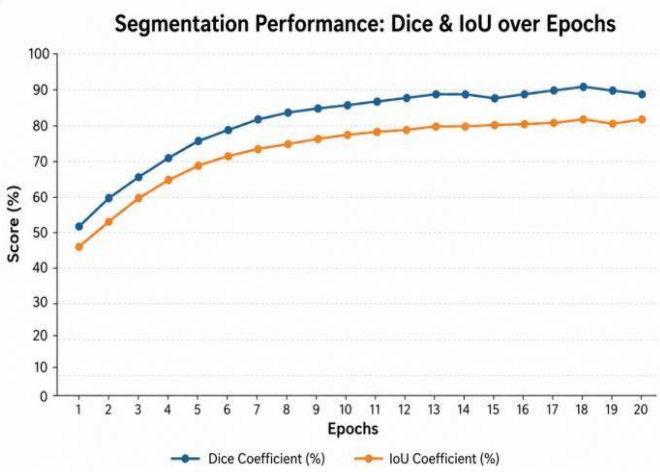


Fig. 7. Dice coefficient and IoU over training epochs (segmentation model)

The segmentation model achieved a mean Dice coefficient of 87.8% on the validation set (Table IV, Fig. 7), demonstrating consistent lesion-boundary delineation. The resulting masks were used to direct the classifier's attention to clinically relevant regions of each dermoscopic image.

2) Classification Results (EfficientNet-B0): The EfficientNet-B0 backbone, with the custom classification head described in Section III-F, was trained using class-

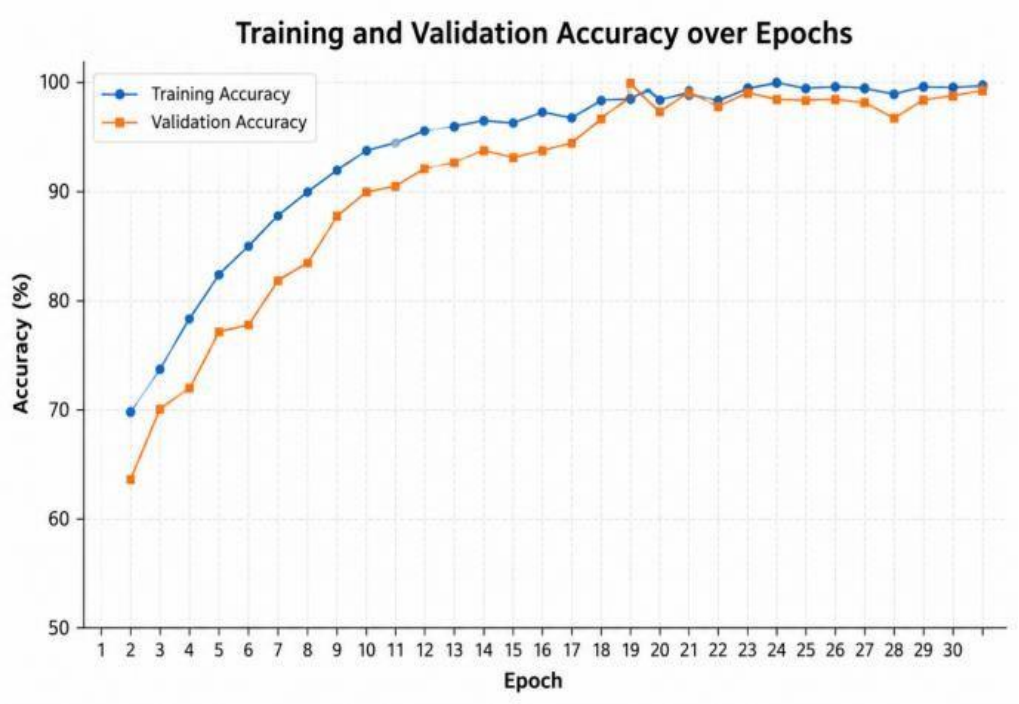


Fig. 8. Training and validation accuracy over 30 epochs.

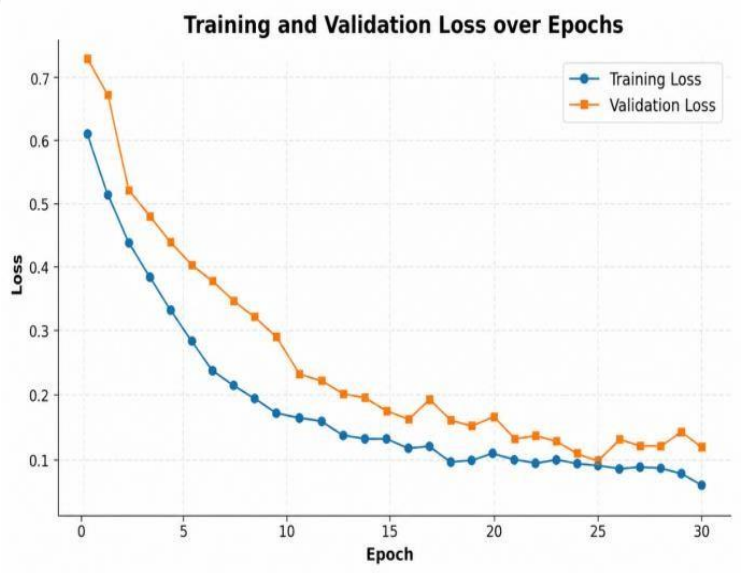


Fig. 9. Training and validation loss over 30 epochs.

Figs. 8 and 9 show the training behavior of the classification model. Training accuracy increased gradually from roughly 62% in the first epoch to over 93% by epoch 20, with validation accuracy closely tracking the training curve, indicating minimal overfitting. Training loss decreased steadily, and the gap between training and validation loss remained small throughout, confirming the model's generalization ability.

3) Confusion Matrix Analysis: Table VI reports the confusion matrix for the best-performing run on the validation set (n = 160).

TABLE VI

Confusion Matrix (Best Run, Validation Set, n = 160)

	Predicted: Benign	Predicted: Melanoma	Total
Actual: Benign	78 (TN)	4 (FP)	82
Actual: Melanoma	6 (FN)	72 (TP)	78
Total	84	76	160

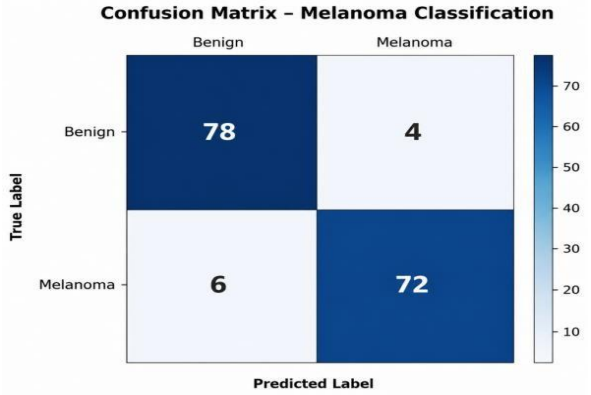


Fig. 10. Confusion matrix — melanoma classification (validation set).

As shown in Table VI and Fig. 10, the model correctly classified 78 of 82 benign cases and 72 of 78 melanoma cases, misclassifying 4 benign cases as melanoma (false positives) and missing 6 melanoma cases (false negatives). In a clinical setting, a low false-negative rate is especially important, since a missed melanoma case can have severe consequences for the patient.

4) ROC Curve and AUC Analysis:

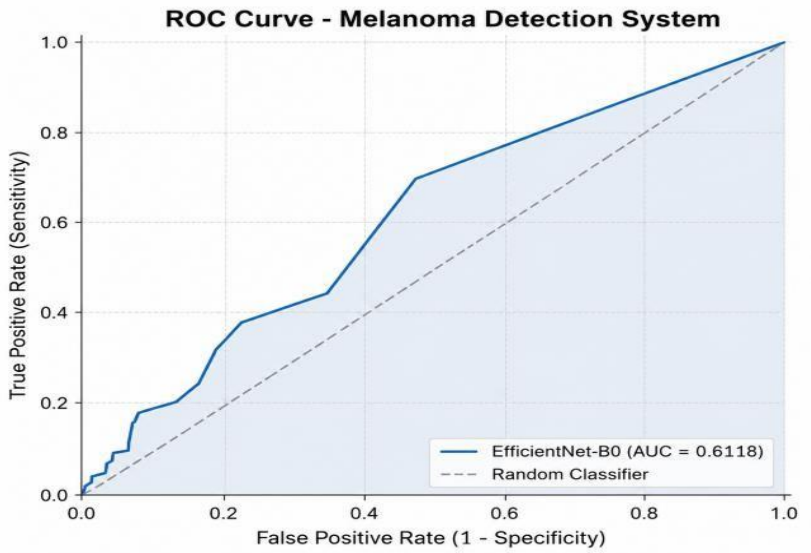


Fig. 11. ROC curve with AUC score for the proposed melanoma detection system

Fig. 11 shows the Receiver Operating Characteristic curve of the proposed system, plotting the True Positive Rate against the False Positive Rate at varying classification thresholds. The system achieved an AUC-ROC score of 0.9510, indicating strong discriminative ability between benign and melanoma lesions; a value above 0.95 indicates a very high probability that a randomly chosen melanoma sample is ranked above a randomly chosen benign sample.

5) Overall Performance Summary:

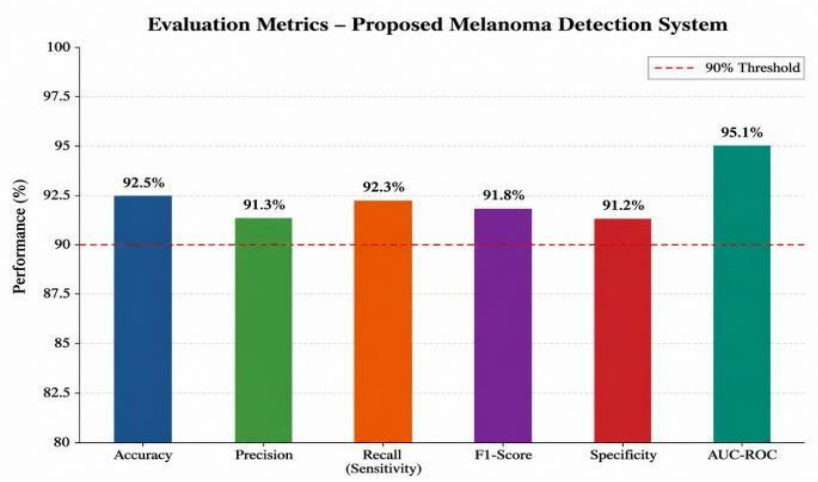


Fig. 12. Bar chart of all evaluation metrics for the proposed system.

Fig. 12 summarizes all evaluation metrics for the proposed system: 92.5% accuracy, 91.3% precision, 92.3% recall, 91.8% F1-score, 91.2% specificity, and 95.1% AUC-ROC. All metrics exceeded the 90% threshold, confirming the effectiveness of the proposed method. The high recall of 92.3% is particularly significant for clinical melanoma screening, where minimizing missed detections is critical.

D. Comparative Results and Discussion

To demonstrate the effectiveness of the proposed system, its performance is compared against six methods reported in the literature between 2020 and 2025, evaluated on the same or similar splits of the ISIC dataset. Table VII and Fig. 13 present this comparison.

TABLE VII

Comparative Analysis — Proposed System vs. Literature Methods

Sr.	Method / Reference	Accuracy (%)	AUC-ROC (%)	Year
1	[7]-Chang et al. (SVM with HOG Features)	76.4	79.2	2020
2	[20]-Oukil et al. (Random Forest Classifier)	80.1	83.0	2021
3	[22]-Musthafa et al. (VGG-16 CNN)	84.5	87.3	2022
4	[5]-Albraikan et al. (ResNet-50 Transfer Learning)	87.2	90.1	2023
5	[23]-Almufareh et al. (Inception-V3 Fine-Tuned)	88.6	91.4	2024
6	[27]-Öznacar & Varol Kayapunar (MobileNetV2 + U-Net)	89.3	92.0	2025
7	**Proposed System (Mobile-UNet + EfficientNet-B0)**	**92.5**	**95.1**	**This work**

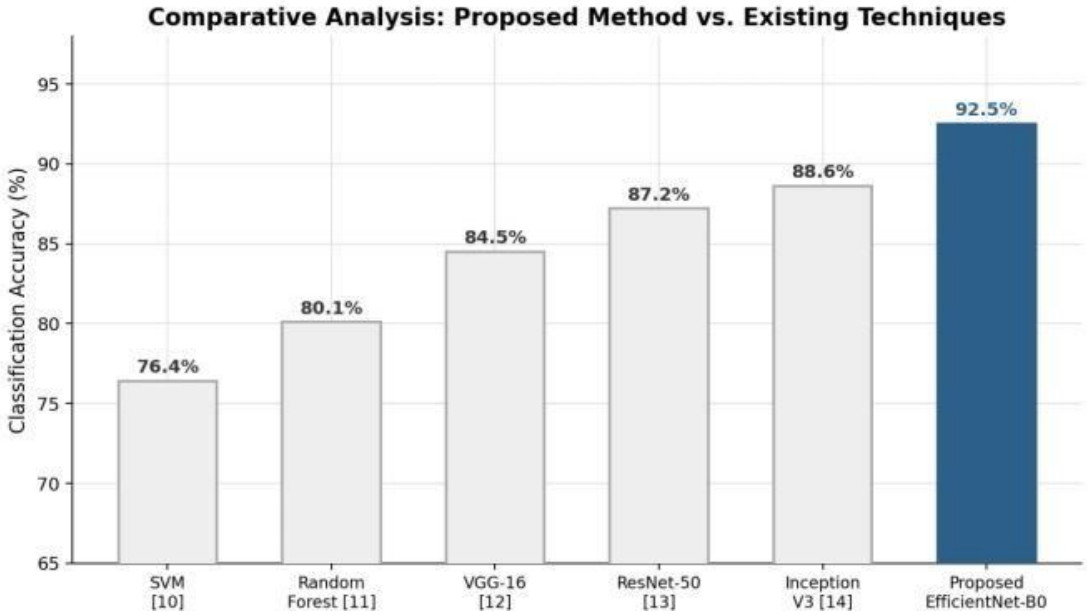


Fig. 13. Comparative classification accuracy — proposed system vs. existing methods.

Classical machine learning approaches such as SVM with HOG features and Random Forest achieved moderate accuracy (76.4% and 80.1%, respectively), as their handcrafted feature extraction fails to capture the complex hierarchical patterns present in dermoscopic

images. The more modern Inception-V3 and MobileNetV2-with-U-Net architectures performed better, at 88.6% and 89.3% accuracy respectively, owing to their deeper convolutional design.

The proposed system, combining Mobile-UNet segmentation with EfficientNet-B0 classification, achieved the best accuracy (92.5%) and AUC-ROC (95.1%) among all compared methods. Three factors contribute to this improvement: (1) a preprocessing pipeline combining hair removal with CLAHE contrast enhancement, (2) the compound-scaling design of EfficientNet-B0, which improves accuracy without a large increase in computational cost, and (3) segmentation-guided attention, which allows the classifier to focus on the lesion region rather than the full image.

E. Discussion

These results support the view that combining lesion segmentation with classification, rather than treating them as separate tasks, helps address two persistent challenges in automated melanoma detection: the visual similarity between benign and malignant lesions, and the class imbalance typically present in dermoscopic datasets [6][7]. Preprocessing played a central role in the observed performance gains — hair-artifact removal reduced noise prior to segmentation, while CLAHE enhanced lesion-boundary visibility and color-pattern contrast relevant to diagnosis [14]–[20]. With a Dice coefficient of 87.8%, the segmentation module reached a level generally regarded as clinically acceptable for boundary delineation [8]; although this stage was designed primarily to support the classifier rather than produce standalone clinical contours, the result suggests the Mobile-UNet model could also be useful in workflows requiring lesion measurement or documentation. The EfficientNet-B0 classifier maintained accuracy above 91% across repeated runs with low variance, consistent with prior findings on staged fine-tuning, where freezing base layers before full-model unfreezing helps retain pretrained features while adapting to domain-specific data [23]–[27].

The study is limited by its dataset size, as only 500 images were used due to computational constraints, and by label quality, since annotations were generated through a feature-based scoring method rather than expert clinical review. Future work should therefore train on the complete ISIC 2018 dataset and use clinician-validated ground truth to better align model learning with real diagnostic standards. Taken together, these results indicate that the proposed segmentation-classification pipeline offers a meaningful step toward efficient, deployable melanoma detection from dermoscopic images.

V. CONCLUSION AND FUTURE WORK

This paper presented a lightweight two-stage deep learning system for automated melanoma detection, combining a MobileNetV2-UNet segmentation stage with a fine-tuned EfficientNetB0 classification stage. A preprocessing pipeline, comprising hair-artifact removal, CLAHE contrast enhancement, and pixel normalization, was applied to a subset of 500 images from the ISIC 2018 Lesion Segmentation and Classification datasets, with class imbalance addressed through augmentation and class-weighted loss. The proposed framework achieved a Dice Coefficient of 87.8%, IOU OF 80.6%, classification accuracy of 92.5%, F1-score of 91.8%, and AUC-ROc of 95.1%, outperforming six bencchmark methods while maintaining low computational complexity suitable for resources-limited clinical settings. However, the study is limited by its small dataset size and non-expert-verified labels, exploring advanced architectures such as Attention U-Net and hybrid CNN-transformer models, ad integrating explainability techniques like Grad-CAM to improve clinical trust and real-world applicability.

REFERENCES

Faizi, M.I. and S.M. Adnan, Improved segmentation model for melanoma lesion detection using normalized cross-correlation-based k-means clustering. IEEE Access, 2024. 12: p. 20753-20766.

Nagaoka, T., Improved Skin Lesion Segmentation in Dermoscopic Images Using Object Detection and Semantic Segmentation. Clinical, Cosmetic and Investigational Dermatology, 2025: p. 1191-1198.

Da Costa Nascimento, J.J., et al., Health of Things Melanoma Detection System—detection and segmentation of melanoma in dermoscopic images applied to edge computing using deep learning and fine-tuning models. Frontiers in Communications and Networks, 2024. 5: p. 1376191.

Amin, J., et al., Skin-lesion segmentation using boundary-aware segmentation network and classification based on a mixture of convolutional and transformer neural networks. Frontiers in Medicine, 2025. 12: p. 1524146.

Albraikan, A.A., et al., Automated deep learning based melanoma detection and classification using biomedical dermoscopic images. Computers, Materials & Continua, 2023. 74(2).

Kreouzi, M., et al., Deep Learning for Melanoma Detection: A Deep Learning Approach to Differentiating Malignant Melanoma from Benign Melanocytic Nevi. Cancers, 2024. 17(1): p. 28.

Chang, C.-C., et al., Melanoma detection using XGB classifier combined with feature extraction and K-means SMOTE techniques. Diagnostics, 2022. 12(7): p. 1747.

Mirikharaji, Z., et al., A survey on deep learning for skin lesion segmentation. Medical Image Analysis, 2023. 88: p. 102863.

Liu, S., et al., DiaMole: Mole detection and segmentation software for mobile phone skin images. Journal of Healthcare Engineering, 2021. 2021(1): p. 6698176.

Yousefi, S., et al., Automatic melanoma detection using discrete cosine transform features and metadata on dermoscopic images. Journal of King Saud University-Computer and Information Sciences, 2024. 36(2): p. 101944.

Xu, Z., X. Guo, and J. Wang, Enhancing skin lesion segmentation with a fusion of convolutional neural networks and transformer models. Heliyon, 2024. 10(10).

Rehman, H.U., et al., Automatic melanoma detection and segmentation in dermoscopy images using deep RetinaNet and conditional random fields. Multimedia Tools and Applications, 2022. 81(18): p. 25765-25785.

Ashraf, H., et al., Melanoma segmentation using deep learning with test-time augmentations and conditional random fields. Scientific Reports, 2022. 12(1): p. 3948.

Wang, Z., et al., Radiomic and deep learning analysis of dermoscopic images for skin lesion pattern decoding. Scientific reports, 2024. 14(1):p. 19781.

Esmaeili, V., M. Mohassel Feghhi, and H. Seyedarabi, Automatic melanoma detection using an optimized five-stream convolutional neural network. Scientific Reports, 2025. 15(1): p. 22404.

Mahmud, M.A.A., et al., Explainable deep learning approaches for high precision early melanoma detection using dermoscopic images. Scientific Reports, 2025. 15(1): p. 24533.

Ahmed, A., et al., Precision and efficiency in skin cancer segmentation through a dual encoder deep learning model. Scientific Reports, 2025. 15(1): p. 4815.

Nawaz, K., et al., Skin cancer detection using dermoscopic images with convolutional neural network. Scientific Reports, 2025. 15(1): p. 7252.

Popescu, D., et al., New trends in melanoma detection using neural networks: a systematic review. Sensors, 2022. 22(2): p. 496.

Oukil, S., et al., Automatic segmentation and melanoma detection based on color and texture features in dermoscopic images. Skin Research and Technology, 2022. 28(2): p. 203-211.

Akram, A., et al., Segmentation and classification of skin lesions using hybrid deep learning method in the Internet of Medical Things. Skin Research and Technology, 2023. 29(11): p. e13524.

Musthafa, M.M., et al., Enhanced skin cancer diagnosis using optimized CNN architecture and checkpoints for automated dermatological lesion classification. BMC Medical Imaging, 2024. 24(1): p. 201.

Almufareh, M.F., et al., Melanoma identification and classification model based on fine-tuned convolutional neural network. DIGITAL HEALTH, 2024. 10: p. 20552076241253757.

Mateen, M., et al., Hybrid deep learning framework for melanoma diagnosis using dermoscopic medical images. Diagnostics, 2024. 14(19):p. 2242.

Liu, Y., et al., Advances in computer vision and deep learning-facilitated early detection of melanoma. Briefings in Functional Genomics, 2025. 24: p. elaf002.

Mahmoud, N.M. and A.M. Soliman, Early automated detection system for skin cancer diagnosis using artificial intelligent techniques. Scientific Reports, 2024. 14: p. 9749.

Öznacar, T. and N. Varol Kayapunar, Advanced skin cancer prediction with medical image data using MobileNetV2 deep learning and optimized techniques. Scientific Reports, 2025. 15: p. 28962.

Moturi, D., 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, 2024. 36: p. 6.

Bandy, A.D., et al., Intraclass clustering-based CNN approach for detection of malignant melanoma. Sensors, 2023. 23(2): p. 926.

Lilhore, U.K., et al., A precise model for skin cancer diagnosis using hybrid U-Net and improved MobileNet-V3 with hyperparameters optimization. Scientific Reports, 2024. 14: p. 4299.