Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 3, July 2026, pp: 1093-1116 e-ISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

Skin Cancer Classification by Applying Different Models of Artificial Intelligence

Wajid Dawood Alwan, Osama Qasim Jumah Al-Thahab, and Hanaa M. Al Abboodi

Department of Electrical Engineering, University of Babylon, Hillah, Iraq

Corresponding author: Wajid Dawood Alwan (e-mail: eng162.wajed.dawood@student.uobabylon.edu.iq), **Author(s) Email**: Osama Qasim Jumah Al-Thahab (e-mail: eng.osama.qasim@uobabylon.edu.iq), Hanaa Mohsin Ali Al Abboodi (e-mail: hanaa.ali@uobabylon.edu.iq)

Abstract **Accurate multiclass classification of dermoscopic skin lesions remains challenging because of high inter-class visual similarity, substantial intra-class variability, and frequent acquisition artifacts (black borders, hair occlusions, noise). We propose a unified, reproducible framework that systematically coordinates four stages: (i) artifact-aware preprocessing (field-of-view circular cropping, hair removal, CLAHE, bilateral filtering); (ii) lesion-focused segmentation via GrabCut-refined fusion and a U-Net with EfficientNet-B3 encoder; (iii) compact deep-feature extraction (EfficientNet-B7) refined by principal component analysis and Neural Spline Flow density calibration; and (iv) robust machine-learning classification. The HAM10000 dataset (n = 10,015, seven diagnostic classes) was partitioned once by stratified random sampling into training (70 %, n = 7010), validation (15 %, n = 1502), and test (15 %, n = 1503) subsets under a strictly sequential anti-leakage protocol with patient-level isolation; the test set was sequestered until terminal evaluation. External generalization was assessed on an independent ISIC 2019 subset (n = 350, 50 per class) without retraining. On the held-out HAM10000 test set, XGBoost achieved the highest accuracy of 99.47 % with an F1-score of 98.99 %, followed by LightGBM (98.20 %) and MLP (97.67 %). Ablation analysis confirmed incremental gains of +2.55 % (preprocessing), +1.75 % (segmentation), and +1.32 % (Neural Spline Flow refinement). On the external ISIC 2019 data, MLP attained the best cross-domain accuracy of 95.43 %, demonstrating that the feature backbone generalizes beyond the training distribution. The demonstrated synergy of artifact suppression, lesion-centered segmentation, and density-calibrated feature learning yields highly discriminative and generalizable representations, providing a robust foundation for reliable computer-aided dermatologic screening.**

Keywords **Bilateral Filtering, Contrast Limited Adaptive Histogram Equalization (CLAHE), Spatial and Channel Squeeze and Excitation (SCSE), Extreme Gradient Boosting (XGBoost), Dermatofibroma, Vascular Lesions.**

I. Introduction

Skin is the body's primary protective barrier; uncontrolled skin-cell proliferation can lead to skin cancer, which primarily affects exposed skin but may also involve areas such as the eyes, nose, and neck [1]. For optimal treatment, early and accurate classification of skin cancer is essential. However, visual inspection and biopsies which are frequently invasive, time-consuming, and operator-dependent are the foundation of traditional diagnosis. These difficulties emphasize the necessity of quicker, more accurate, and more trustworthy diagnostic methods. Globally, skin cancer is a serious health issue. Medical professionals detect an estimated 2-3 million non-melanoma skin cancers and 130,000 melanoma skin cancers each year. Due to its rapid growth and ability to spread to other parts of the body, melanoma is the least frequent but most dangerous type of skin cancer. Because it can raise the 5-year mortality rate to about 99%, early diagnosis of malignant tumors is crucial [2]. Accordingly, the HAM10000 dataset is widely adopted in this domain and comprises seven lesion types: Actinic keratosis, Basal cell carcinoma, Benign keratosis-like lesions, Dermatofibroma, Melanocytic nevi, and Vascular lesions [2].

Automated skin cancer analysis has undergone a revolution thanks to major advances in machine learning (ML) and deep learning (DL) over the past few years. Brain-inspired neural architectures that use vast volumes of labeled data to identify representative patterns have recently demonstrated tremendous promise in AI [3], [4]. The primary area where convolutional neural networks (CNNs) excel in medical image processing is their high performance, which can incur high computational cost, even though computer-aided systems can help with the diagnostic process. Algorithms commonly used to enhance diagnostic performance and streamline clinical procedures include support vector machines, decision tree classifiers, and convolutional neural networks [5]. For optimal treatment, early detection of skin cancer is crucial. While CNNs may be trained to directly extract

Manuscript received 14 May 2026; Revised 22 June 2026; Accepted 28 June 2026; Available online 6 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1757

high-level features from dermoscopic images, earlier work used ML classification and feature extraction/selection steps independently, leading to complex models with a high risk of overfitting [6], [7].

Recent HAM10000 classifiers can be synthesized into four methodological paradigms. First, hybrid feature-based pipelines are represented by Mostafiz et al. [8], who combine conventional classifiers with morphological black-hat filtering and inpainting to achieve 97% accuracy, and by Hritwik et al. [17], who reach 98.75% training accuracy with a hybrid DL model on 3,000 images; both decouple feature extraction from classification, propagating redundant or noisy representations without systematic dimensionality reduction or density calibration. Second, end-to-end CNNs include Ziyi Li et al. [9] (quantum Inception-ResNetV1, 98%), Nigar et al. [11] (novel deep model, 97%), Vipin et al. [12] (modified EfficientNet, 95.49%), Marwa et al. [13] (IoT-optimized algorithm, 98.66%), Muhammad Azeem et al. [14] (SkinLesNet, 96%), Mohamad et al. [16] (FT-DenseNet201, 100% training metrics), Rafik Ahmad et al. [19] (ResNet-50, >92%), S. Gorgbandi and S. Nazari [20] (CAD deep CNNs, 90.5%), and P. Chaudhary [21] (AI early detection, 92%); these achieve high accuracy yet operate as opaque black boxes prone to overfitting, without systematic artifact suppression, validated lesion-centered segmentation, or independent external validation. Third, segmentation-assisted methods comprise Aishwarya et al. [10] (YOLOv3, 88.03% mAP), Vankayalapati Radhika and B. Sai Chandana [15] (DenseNet-U-Net links, ISIC 2019), and Razia et al. [18] (98.345% without segmentation versus 97.757% with segmentation), demonstrating that naive masking can inadvertently degrade performance by contaminating feature learning rather than refining it. Fourth, transformer-based architectures are not represented among the cited HAM10000 benchmark studies [8] to [21], which predominantly reflect CNN and hybrid paradigms; this absence underscores the field's concentration on convolutional approaches and the lack of unified frameworks that transcend isolated module optimization. Consequently, none of the existing studs among [8] to [21] coordinates artifact-aware preprocessing, lesion-centered segmentation with validated pseudo-ground-truth, compact deep-feature refinement via PCA and Neural Spline Flow density calibration, and robust cross-domain evaluation. The proposed pipeline closes this gap by synergistically integrating these four stages, achieving 99.47% internal accuracy and 95.43% cross-domain generalization.

Despite recent improvements in HAM10000 categorization, there is still a significant fragmentation among current approaches. Classical machine-learning pipelines function without systematic refinement of representations and propagate redundant or noisy representations, in contrast to end-to-end deep models, which are opaque, black-box-type models with high computational costs that may overfit on limited dermoscopic data. Black borders, hair occlusion, artifact corruption, and seemingly corrected acquisition noise are examples of non-lesion structures that taint downstream learning. In order to maintain border irregularities and color variegation, diagnostic analysis is typically not combined with segmentation; dimensionality reduction and density calibration are infrequently employed to improve feature reliability. Crucially, aside from one study, there was minimal external validation across separate datasets to support generalizations. Crucially, aside from one study, there was insufficient external validation across other datasets to support generalizations. Consequently, no unified framework systematically coordinates artifact-aware preprocessing, lesion-centered segmentation, compact deep feature refinement via PCA and Neural Spline Flow density calibration, and robust machine-learning classification with independent external validation. This gap necessitates the proposed pipeline.

To address this gap, this paper presents a reproducible classification framework for HAM10000 that integrates artifact-aware preprocessing, lesion-focused segmentation, compact deep feature extraction, and machine-learning prediction within a unified pipeline. Preprocessing removes black-border artifacts through field-of-view-based circular ROI detection and suppresses hairs using morphological black-hat filtering followed by Telea inpainting; bilateral filtering and Contrast Limited Adaptive Histogram Equalization (CLAHE) are then applied to reduce noise and enhance local contrast. Lesion-focused analysis is performed using a GrabCut-refined color–texture fusion scheme with distance-transform core cropping. Meanwhile, lesion masks are generated via Otsu thresholding and Lab-space K-means clustering, and these masks supervise a U-Net model with an EfficientNet-B3 encoder and Spatial and Channel Squeeze-and-Excitation (SCSE) attention, trained for 65 epochs using a composite Dice–Binary Cross-Entropy (BCE)–Focal loss. To improve robustness and class balance, rotation, shift, flip, zoom, and brightness augmentation are applied to the original images. High-level lesion descriptors are extracted using EfficientNet-B7 and subsequently refined by principal component analysis and Neural Spline Flow (NSF) before classification with Random Forest (RF), Histogram Gradient Boosting (HistGradient Boosting), Extreme Gradient Boosting (XGBoost), LightGBM, Tabular Network (TabNet), and Multi-Layer Perceptron (MLP) models. The principal contributions of this study are fourfold and explicitly differentiated from existing

Manuscript received 14 May 2026; Revised 22 June 2026; Accepted 28 June 2026; Available online 6 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1757

HAM10000 classification literature. First, we propose an artifact-aware preprocessing pipeline that unifies field-of-view-based circular ROI detection with inscribed 4:3 rectangle cropping, morphological black-hat filtering with Telea inpainting for hair removal, and an adaptive CLAHE–bilateral filtering protocol governed by Laplacian sharpness metrics. Second, we introduce a lesion-focused segmentation framework combining a GrabCut-refined color–texture fusion scheme with distance-transform core cropping, along with a U-Net architecture with an EfficientNet-B3 encoder and spatial-channel squeeze-and-excitation (SCSE) attention trained via a composite Dice BCE–Focal loss. Third, we present a compact feature refinement strategy in which 2560-dimensional EfficientNet-B7 descriptors are reduced by principal component analysis, with the truncation level determined through a two-stage coarse-to-fine search using a stacking ensemble and subsequently refined by Neural Spline Flow (NSF) to enhance discriminative reliability through calibrated density estimation. Fourth, we rigorously compare six machine-learning classifiers on the refined feature space and externally validate the framework on the ISIC 2019 dataset, achieving 99.47% accuracy on HAM10000 and up to 95.43% on external data. The findings demonstrate that performance gains are obtained by methodically integrating various modules, such as compact deep-feature learning, lesion-centered segmentation, strong ensemble classification, and artifact suppression.

The properties of the data, such as data balancing approaches, segmentation, deep feature extraction, dimensionality reduction, and creation of the hybrid classification models, are thoroughly covered in Section II along with the suggested methodological framework. The performance criteria and evaluation indicators used to quantify a model's effectiveness are outlined in Section III. The experimental results are explained in Section IV. Section V contains the discussion. Section VI explains comparative analysis. Finally, Section VII summarizes the results and discusses the research's implications and future directions.

II. Method

The proposed pipeline Fig. 1 comprises six sequential stages: (i) artifact-aware preprocessing; (ii) GrabCut-refined lesion segmentation via U-Net with EfficientNet-B3 encoder and SCSE attention; (iii) ROI cropping and geometric-photometric augmentation; (iv) 2560-dimensional feature extraction with EfficientNet-B7; (v) PCA-NSF compact feature refinement; and (vi) classification using HistGBM, XGBoost, RF, LightGBM, MLP, or TabNet. The dataset was stratified into training (70%, n = 7010), validation (15%, n = 1502), and test (15%, n = 1503) subsets under a strictly sequential anti-leakage protocol with patient-level isolation. Performance was assessed using accuracy, precision, recall, F1-score, and specificity.

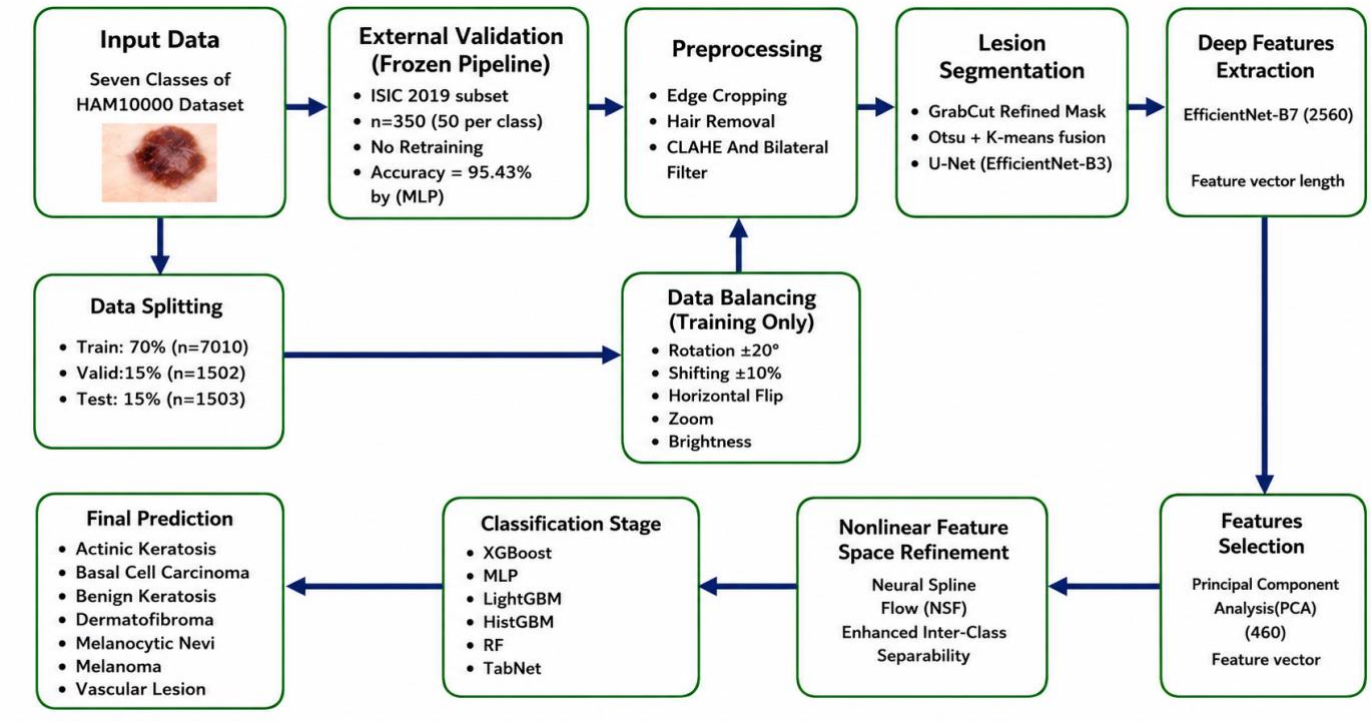


Fig. 1. Proposed pipeline for seven-class skin lesion classification.

Manuscript received 14 May 2026; Revised 22 June 2026; Accepted 28 June 2026; Available online 6 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1757

Strictly sequential data-balancing and anti-leakage protocol. To ensure reproducibility and prevent information leakage, the HAM10000 dataset was partitioned once, before any processing, into training (70%, n = 7010), validation (15%, n = 1502), and test (15%, n = 1503) subsets via stratified random sampling at the image level with a fixed random seed. In rare cases where multiple images correspond to the same lesion, all such images were confined to a single partition. The test subset was physically sequestered and used only for the terminal evaluation reported in Table 9 and Figs. 9–12. Preprocessing, segmentation, augmentation, PCA, NSF, and classifier training were executed exclusively on the training set; the validation set was employed solely for early stopping and hyperparameter selection. No test-set image, label, or statistic was exposed to any model-fitting or model-selection decision.

A. Dataset preparation

This work extensively uses the HAM10000 dermoscopic image dataset, obtained from Kaggle [22]. The dataset contains 10,015 images spanning seven diagnostic categories: Actinic Keratoses (AK), Basal Cell carcinoma (BCC), Benign Keratosis-Like Lesions (BKL), Dermatofibroma (DF), Melanocytic Nevi (NV), Melanoma (MEL), and Vascular Lesions (VASC). The images were gathered from a range of clinical sources, including the Department of Dermatology at the Medical University of Vienna and the Skin Cancer Practice of Cliff Rosendahl in Queensland, Australia, due to the clinical variability and expert-verified diagnostic labeling.

Table 1. HAM10000 dataset distribution

Category	Abbreviation	Number of images
Actinic keratosis	AK	327
Basal Cell Carcinoma	BCC	514
Benign Keratosis-like Lesions	BKL	1099
Dermatofibroma	DF	115
Melanocytic Nevi	NV	6705
Melanoma	MEL	1113

Because of its size, diversity, and consistent annotations, HAM10000 is widely used as a benchmark set for computer-aided diagnostic methods for dermoscopic image analysis. Representative samples from each class are illustrated in Fig. 2, while the class-wise distribution is summarized in Table 1.

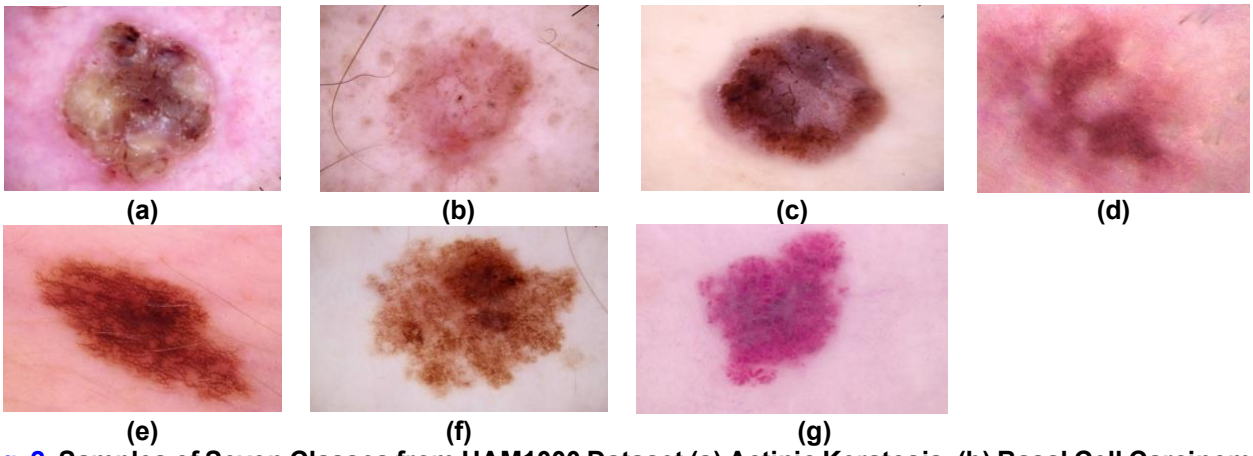


Fig. 2. Samples of Seven Classes from HAM1000 Dataset (a) Actinic Keratosis, (b) Basal Cell Carcinoma, (c) Benign Keratosis, (d) Dermatofibroma, (e) Melanocytic Nevi, (f) Melanoma, (g) Vascular Lesion.

To avoid information leakage and guarantee data reproducibility, the HAM10000 dataset was divided into three subsets: training (70% of the cases, n = 7010), validation (15% of the cases, n = 1502), and test (15% of the cases, n = 1503) using stratified random sampling at the lesion level with lesion_id metadata as the stratification variable. After that, every image with the same lesion_id was placed in the same partition and therefore considered to belong to the same patient. The test subset was physically sequestered and used only for terminal evaluation. Preprocessing, segmentation, augmentation, PCA, NSF, and classifier training were executed exclusively on the training set; the validation set was employed solely for early stopping and hyperparameter selection. No test-set image, label, or statistic was exposed to any model-fitting or model-selection decision.

B. Pre-processing

The input images were preprocessed and optimized to support skin-disease classification, enhancing model generalizability and reliability through the following steps: black-border cropping, hair-removal artifact

Manuscript received 14 May 2026; Revised 22 June 2026; Accepted 28 June 2026; Available online 6 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1757

suppression, CLAHE-based local contrast enhancement, and edge-preserving denoising using bilateral filtering.

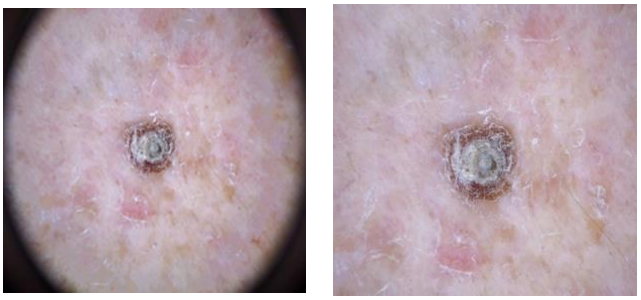
1. Black border cropping.

Non-informative black-and-white edges are removed using FOV-based circular ROI detection, followed by inscribed 4:3 rectangle cropping. Images are converted to HSV; a valid-region mask is estimated by thresholding dark and low-saturation bright pixels, then morphologically cleaned. The connected valid component centered on the image is approximated by a minimum enclosing circle (Hough-circle fallback if needed). The largest inscribed 4:3 rectangle is cropped with a small inward margin, post-trimmed, and standardized to 600×450 via reflection padding. Fig. 3 shows the image before and after processing. Let W and H denote the image width and height, respectively, and define the aspect ratio a = H/W. Given the effective radius r′ after margin adjustment, the half-height h and half-width w of the inscribed 4:3 rectangle are computed in Eq. (1) and Eq. (2) [23] as:

$$h = \frac{r'}{\sqrt{1+a^2}} \quad (1)$$

$$w = \frac{ar'}{\sqrt{1+a^2}} \quad (2)$$

The crop spans [cx − w, cx + w] and [cy − h, cy + h] are then centered at the detected circle center (cx, cy).

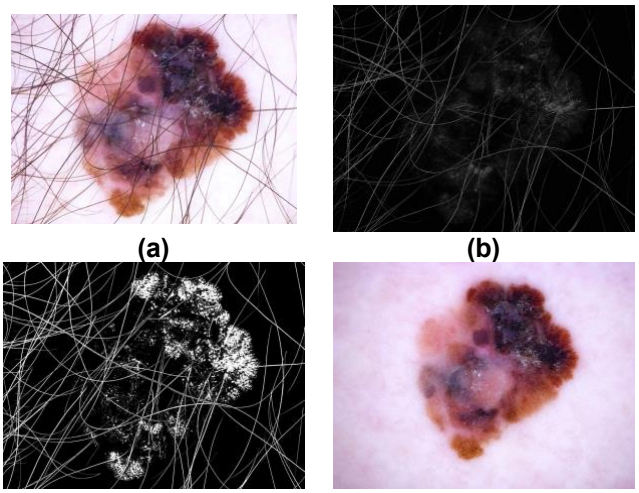


(a) (b)

Fig. 3. Sample of Skin cancer (a) Original Image (224,224,3), (b) image after black edge removal.

2. Hair removal

A two-stage pipeline that applies Telea's inpainting for the restoration process and morphological black-hat filtering for hair localization implements automated hair artifact suppression. To improve the contrast between the skin backdrop and the thin dark hairs, all dermoscopic images were first converted to grayscale. The image's hair-like characteristics were then highlighted and made more noticeable using a black-hat morphological operator. A binary mask with background pixels set to 0 and hair pixels set to 1 was created by thresholding the response. The Telea inpainting technique, which propagates information from nearby unoccluded skin pixels to the positions of the hair, was then used to fill in the masked areas. As seen in Fig. 4 [24], [25], this produces a smooth appearance of the restored areas and visually consistent merging with the surrounding tissue.



(a) (b) (c) (d)

Fig. 4. Image produced after hair removal technique (a) Original Image (224,224,3), (b) Black- hat filter, (c) Binary Mask of Hair, (d) Hair removal.

3. CLAHE and Bilateral Filter

Adaptive control was employed together with a bilateral filter to denoise while maintaining edge structure. The Laplacian variance was used to estimate image sharpness, and it was reduced when sharpness was too low to prevent excessive smoothing [10]. The local contrast of the luminance-related component, the V channel of HSV or the L channel of CIE (International Commission on Illumination) Lab, was then improved using Contrast Limited Adaptive Histogram Equalization (CLAHE), which enhances illumination uniformity and fine detail visibility. To avoid under- and over-amplification of the image, the improved output was linearly mixed with the original image in a ratio of 60% improved and 40% original. Finally, an unsharp masking step was conditionally activated when the measured sharpness fell below a predefined threshold to restore perceived detail. The overall filtering workflow is illustrated in Fig. 5 [26].

C. Dataset segmentation

In this work, a GrabCut-refined color texture fusion segmentation with distance-transform core cropping to extract lesion-centered patches from HAM10000 dermoscopic images. In this approach, a continuous lesion likelihood map S(y,x) is first computed by fusing normalized inverse luminance (Lab), normalized inverse brightness (HSV), and Laplacian-based texture/edge energy under a Gaussian spatial prior Eq. (3), then binarized using Otsu's criterion to obtain the initial lesion mask M0 Eq. (4) [27].

Manuscript received 14 May 2026; Revised 22 June 2026; Accepted 28 June 2026; Available online 6 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1757

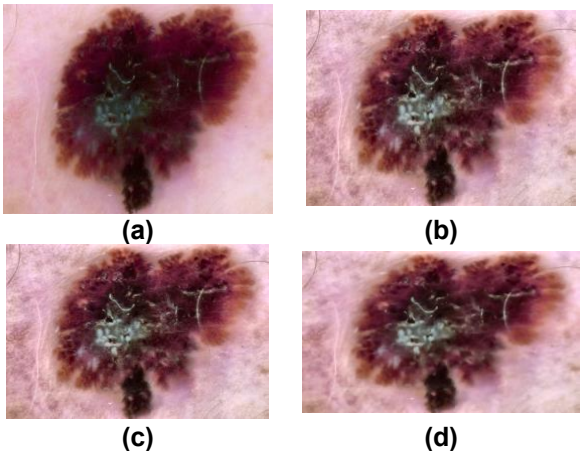


Fig. 5. Sample image from the dataset (a) Original Image, (b) CLAHE filter, (c) Input image to bilateral filter, (d) Output image after bilateral.

$$S(y,x) = (\alpha\, L\_n\,(y,x) + \beta\, V\_n\,(y,x) + \gamma\, E\_n\,(y,x))G(y,x) \quad (3)$$

where S(y,x) is the continuous lesion saliency map, $L_n(y,x)$, $V_n(y,x)$, and $E_n(y,x)$ denote the normalized inverse luminance (Lab), normalized inverse brightness (HSV), and Laplacian-based texture/edge energy, respectively; G(y,x) is a Gaussian spatial prior centered on the lesion; and α, β, and γ are fusion weights. center-prior Gaussian or an adaptive mask (e.g., from a coarse segmentation). Coefficients of α, β, and γ are set to: (α, β, γ) = (0.55,0.20,0.25)

$$M_0(y,x) = 1[S(y,x) >= T^*] \quad (4)$$

where $M_0(y,x)$ is the initial binary lesion mask obtained by thresholding the saliency map,1[ ] denotes the indicator function (equal to 1 if the condition is true and 0 otherwise), and T* is the optimal threshold determined by maximizing the between-class variance via Otsu's criterion, i.e., $T^* = \mathrm{argmax}_T\, \sigma_B^2(T)$.

The mask is subsequently regularized (opening/closing, largest connected component, hole filling) and refined using GrabCut, initialized with confident foreground/background seeds [27]. A conservative core region M core is extracted by thresholding the Euclidean distance transform d(p) at distance d_0, as formulated in Eq. (5). The lesion crop bounding box B′ is then expanded by a proportional margin ρ relative to the core dimensions Δx and Δy, as given in Eq. (6) [22]; the crop is resized to 256×256 using aspect-ratio-preserving padding.

$$M_{core(p)} = 1[d(p) \geq d_0] \quad (5)$$

where $M_{core(p)}$ is the binary core-region mask, $d(p)$ is the Euclidean distance from pixel p to the nearest background pixel computed by a distance transform, and d_0 is the distance threshold that defines the minimum distance required for inclusion in the core. The expanded lesion crop bounding box is given by Eq. (4) [28].

$$B' = (y_0 - \rho\Delta y, y_1 + \rho\Delta y, x_0 - \rho\Delta x, x_1 + \rho\Delta x) \quad (6)$$

where B' denotes the expanded lesion crop bounding box, ρ is the proportional padding ratio, and x_0, x_1, y_0,

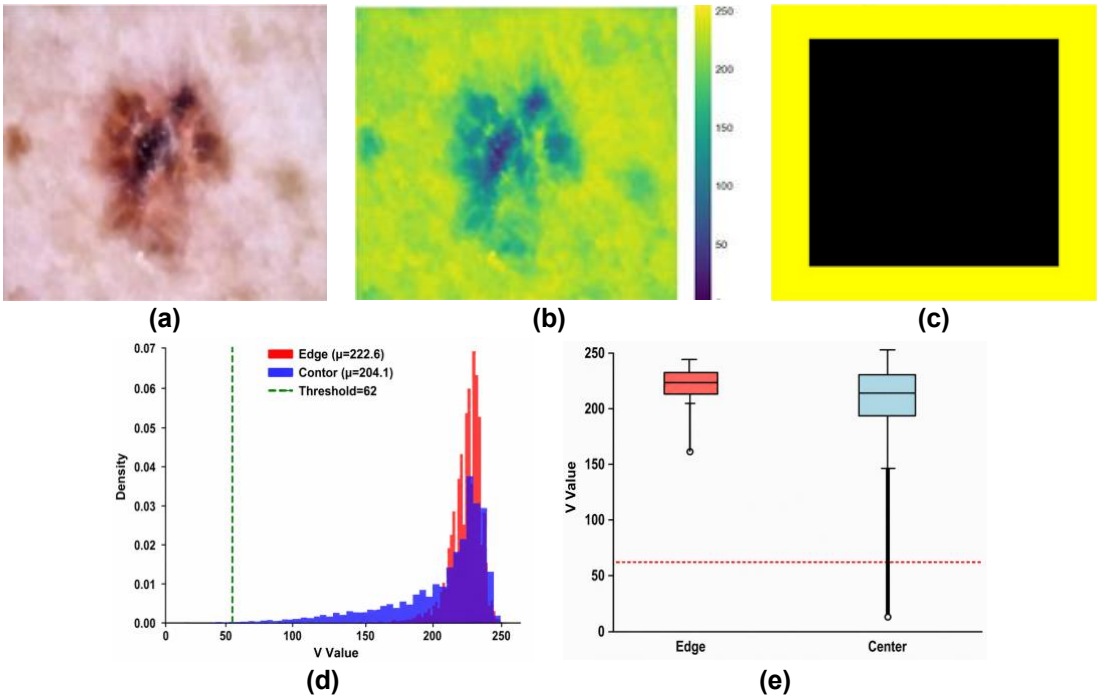


Fig. 6. Ring detection illustration (a) Original image, (b) Ring detection, (c) Edge detection, (d) Intensity distribution, (e) Box plot comparison.

Manuscript received 14 May 2026; Revised 22 June 2026; Accepted 28 June 2026; Available online 6 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1757

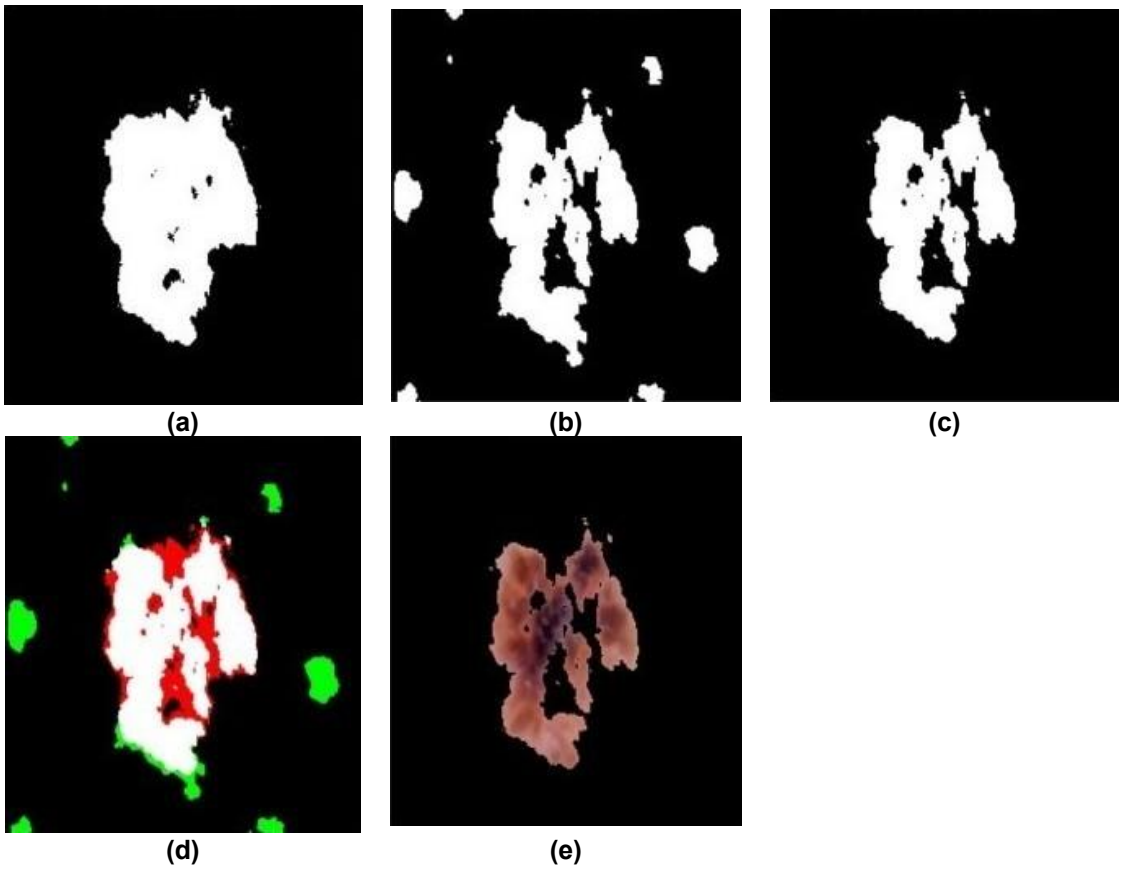


Fig. 7. Mask generation step (a) Otsu Mask, (b) Mask generation, (c) Combined mask, (d) Comparison, (e) Masked image.

y_1 are the top-left and bottom-right corners of the core mask M_{core} , Δx and $\Delta y$are the core width and height [28]. The ring-screening step (edge-band analysis on the HSV V-channel) reports zero low-V edge coverage and therefore classifies the sample as no ring Fig. 6, implying that no ring margin is required before segmentation. The complementary statistical comparison between edge and center V-intensity distributions shows μedge = 222.6 (σ = 9.9) and μcenter = 206.1 (σ =35.8) with Cohen's d = −0.628.

For mask construction, Otsu thresholding and Lab K-means clustering yield masks with coverages 0.173 and 0.165, respectively, while their intersection produces a more conservative combined mask with coverage 0.132 (Fig. 7), reducing scattered background detections; the agreement between the two masks is quantified by Jaccard = 0.644 and pixel agreement = 0.927. (Fig. 10) shows the feature-analysis visualization of luminance L∗, which contributes the largest share of the fused signal (with b∗ and a∗ providing additional discrimination), and the saliency map concentrates its highest responses over the lesion core, supporting that the fusion in Eq. (3) emphasizes diagnostically relevant lesion structures rather than surrounding skin.

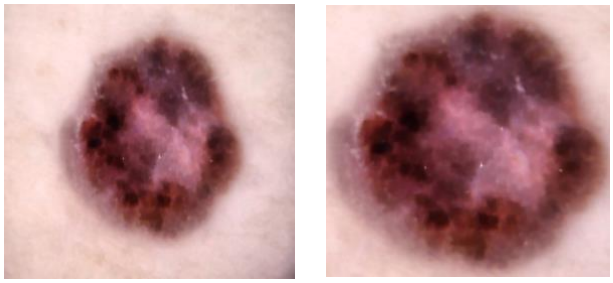


Fig. 8. Sample image from the dataset (a) Original Image, (b) after cropping.

For the HAM10000 dermoscopic dataset, an automated preprocessing pipeline was created as a regular technique to eliminate background artifacts and portray the lesions consistently. A U-Net architecture utilizing an EfficientNet-B3 encoder and a spatial-channel squeeze-and-excitation attention mechanism was trained for 65 epochs with a combined loss that weighted Dice, BCE, and Focal components, following preliminary preprocessing steps such as CL-AHE and bilateral filtering to enhance vascular structures while suppressing acquisition noise. Quantitative segmentation performance is reported independently

Manuscript received 14 May 2026; Revised 22 June 2026; Accepted 28 June 2026; Available online 6 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1757

in Section IV-B using Dice, IoU, sensitivity, and specificity metrics computed on the held-out test set against GrabCut-refined masks.

The trained models produced standardized 256x256-pixel crops of the diseased regions of interest by generating precise binary masks of the lesions and the surrounding healthy integument. As shown in Fig. 8, this segmentation-based cropping technique effectively eliminates undesired structures from the image, such as hair follicles, calibration rulers, and other tissue features, while preserving informative lesion features, such as asymmetry, border irregularity, and color variegation, which are important for diagnosis. This allows the classifications to rely only on tissue patterns that are important for diagnosis.

D. Dataset augmentation and Balancing

A new ImageGenerator class was created to apply stochastic geometric and photometric transformations as part of an extensive multimodal data augmentation procedure in order to address the inherent class imbalance and small sample sizes of dermatological image datasets, notably HAM10000.

Table 2. Deep Feature Extraction using EfficientNet-B7.

Stage	Operator / Block Type	Kernel Size	Stride	Input Image	Output Channels	Feature Map Size
1	Initial Conv	3×3	2	600×600	64	300×300
2	MBConv1	3×3	1	300×300	32	300×300
3	MBConv6	3×3	2	300×300	48	150×150
4	MBConv6	5×5	2	150×150	80	75×75
5	MBConv6	3×3	2	75×75	160	38×38
6	MBConv6	5×5	1	38×38	224	38×38
7	MBConv6	5×5	2	38×38	384	19×19
8	MBConv6	3×3	1	19×19	640	19×19
9	Final Conv 1×1	-	19×19	2560	1×1	

$$W_{new} = \frac{W_{old} + 2P - k}{S} + 1 \quad (7)$$

$$H_{new} = \frac{H_{old} + 2P - k}{S} + 1 \quad (8)$$

Where W_{new} and H_{new} denote the output width and height, W_{old} and H_{old} denote the input width and height , K is the kernel size, S: stride size, and P is the padding size. For large spatial extents and typical stride values, these expressions simplify to $W_{new} \approx W_{old}/S$ and $H_{new} \approx H_{old}/S$. The number of features obtained by efficient B7 is 2560, which are fed to the feature selection stage. Table 2 shows the details of image feature extraction. The network accepts an RGB input with a maximum size of 600×600. Across successive stages, the spatial resolution and channel depth are jointly transformed by convolutional operations (including strided downsampling where applicable). The output spatial dimensions after each such operation are given by Eq. (7), and Eq. (8) [29] , [30]:

To ensure an unbiased and clinically realistic evaluation, the augmentation pipeline described above was applied only to the training split (70 %). To preserve the inherent imbalance observed in clinical practice, the class ratios were maintained in both the validation and test sets. A strictly sequential data-splitting and anti-leakage protocol was enforced: images were first split at the patient level where possible, then augmentation was performed after the training subset was isolated, eliminating any risk of augmented replicas leaking into validation or test folds.

E. Feature extraction and selection

After preprocessing, deep feature vectors were extracted using the Efficient Net family (B0–B7). Table 2 reports the canonical stage-wise configuration of EfficientNet-B7, from the stem convolution to the final convolutional block, as defined in the original architecture specification. As seen in the output visualization of the melanoma class, the augmentation strategy employed is specifically controlled variation: random rotation (between -20° and +20°), width shifting (between -10% and +10%), height shifting (between -10% and +10%), horizontal flipping, zoom (between 0.8× and 1.2×), and multiplicative brightness (between 0.8× and 1.2×) in the HSV color space. In order to preserve the integrity of the lesion boundary and add realistic morphological variations, these modifications are implemented using an affine transformation in OpenCV with reflection padding.

1. Principal Component analysis (PCA)

Feature selection and dimensionality reduction were performed to identify a compact and discriminative representation of the EfficientNet-B7 feature space while minimizing redundancy and improving downstream classification stability. After loading the extracted feature CSV, non-informative identifiers (e.g., reference/image IDs) were removed, class labels were encoded into numeric form, and the remaining feature vectors were converted to floating-point values. The data were then stratified into training and test sets (15% hold-out), and all feature dimensions were standardized using z-score normalization (StandardScaler) to ensure comparable scaling before projection. An orthogonal lower-dimensional basis that

Manuscript received 14 May 2026; Revised 22 June 2026; Accepted 28 June 2026; Available online 6 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1757

maintained the dominant variance structure was produced by using Principal Component Analysis (PCA) on both the training and test partitions, but only on the training data. A series of PCA truncation levels up to 500 components is computed in a coarse search (50 component steps) and a fine-grained search (a ±20 step around the best coarse answer) to find the ideal number of kept components. For every potential dimensionality, n, a stacking ensemble was trained on the training subset and tested on the validation subset. The most informative feature subset was determined by selecting the optimal n that yielded the maximum accuracy on the validation set [31], [32].

Crucially, this is a formal tiered validation process: the stacking ensemble's performance on the validation data alone determines the degree of truncation, and the PCA is only fitted to the training data. This is due to the fact that the held-out test set is used to objectively assess performance and is never utilized in this decision-making process [33].

2. Neural Spline Flow (NSF)

Normalizing flows model complex, high-dimensional distributions by composing invertible transformations with a tractable base distribution pz(z), typically a standard multivariate Gaussian.

The feature vectors with 2560 length are reduced using Principal Component Analysis (PCA) to achieve compact, orthogonal representations while retaining most of the dataset's variance. Let x ∈ RN×460 denote the centered feature matrix, obtained by subtracting the mean vector μ from each feature, $\tilde{X} = X - \mu$. The covariance structure of the data is described by Eq. (9) [34].

$$C = \frac{1}{N-1}\ \tilde{X}^{T}\tilde{X} \tag{9}$$

where C is the sample covariance matrix, N denotes the number of samples, $\tilde{X} = X - \mu$, is the mean-centered data matrix with μ the feature-wise mean vector) and $\tilde{X}^{T}$ denotes its transpose. The product yields the pairwise covariances between all feature dimensions.

The product $\tilde{X}^{T}\tilde{X}$ yields the pairwise covariances between all feature dimensions, thereby capturing linear dependencies in the original space. Principal Component Analysis (PCA) is obtained by eigen-decomposing C,, producing eigenvalue–eigenvector pairs (λ_i, v_i) that satisfy $cv_i = \lambda_i v_i$. The eigenvalues are sorted in descending order so that earlier components correspond to directions of maximal variance. To form a compact representation, the leading m = 460 eigenvectors are stacked to define the projection matrix $W_m = [v_1, v_2, \ldots, v_{460}]$, and the reduced feature embedding is computed in Eq. (10) [34]:

$$Z = \tilde{X}W_m \tag{10}$$

where Z is the reduced feature embedding matrix, $Z \in R^{N\times m}$, where each row is the PCA-transformed feature vector of a sample, W_m is the projection matrix $W_m = [v_1, v_2, \ldots, v_m]$ formed by concatenating the leading m eigenvectors of the sample covariance matrix C,m is the number of retained principal components (m=460 in this work), and the amount of information retained by the selected components is measured through the explained variance ratio as shown in Eq. (11) [34]:

$$Explained\ Variance(460) = \frac{\sum_{i=1}^{460}\lambda i}{\sum_{i=1}^{2560}\lambda i} \tag{11}$$

indicating that the first 460 principal components preserve the dominant variance of the original 2560-dimensional feature space while suppressing redundant correlations and noise. This orthogonal projection yields a more compact yet discriminative representation that is well-suited for subsequent classification, as shown in Eq. (12) [35].

$$P_x(x) = pz\left(f^{-1}(x)\right)\left|\det\frac{\partial f^{-1}(x)}{\partial x}\right| \tag{12}$$

where z(0) = x and z(L) = z denotes the intermediate representations. The Neural Spline Flow (NSF) implements each transformation fl as an affine coupling layer interleaved with random permutation matrices. Given a binary mask m ∈ {0,1}D that partitions the input into fixed components xm = x ⊙ m and transformed components x_m = x ⊙ (1-m), the coupling layer applies the transformation $y_m = x_m$ and y_m = g (x_m ; Θ (x_m) where g denotes a monotonic rational-quadratic spline parameterized by Θ, which are generated by a residual neural network conditioned on the masked components x_m. Within the k-th bin bounded by knots (x_k,y_k) and (x_{k+1},y_{k+1}) , the spline transformation is given by Eq. (13) [35]:

$$y = y_k + \frac{(y_{k+1}-y_k)[s_k(x-x_k)+\delta_k(x-x_k)^2]}{s_k + [\delta_{k+1}+\delta_k - 2s_k](x-x_k) + \delta_k(x-x_k)^2} \tag{13}$$

where $s_k = \frac{y_{k+1}-y_k}{x_{k+1}-x_k}$ is the linear slope between adjacent knots (x_k, y_k) and (x_{k+1} , y_{k+1}), and δk>0 denotes the learned derivative at the k-th knot, ensuring strict monotonicity and analytical invertibility[35].

It is important to delimit the empirical scope of the NSF module herein. While normalizing flows theoretically enable density-based uncertainty estimation, the NSF is employed in this study strictly as a nonlinear feature refinement mechanism that recalibrates PCA-reduced embeddings to enhance inter-class separability. Explicit uncertainty metrics such as expected calibration error, reliability diagrams, or confidence-based rejection thresholds are not reported and are reserved for subsequent clinical-deployment investigations.

F. Classification Models

Although transfer learning is widely effective in natural-image domains, its performance can be constrained in medical image analysis due to domain shift, where the source pretraining distribution differs substantially from the target clinical imaging characteristics (e.g., texture statistics, acquisition conditions, and lesion

Manuscript received 14 May 2026; Revised 22 June 2026; Accepted 28 June 2026; Available online 6 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1757

morphology), potentially limiting feature transferability [36].

In this work, a set of complementary classification back-ends is utilized: Tabular Learning Network (TabNet), Multi-Layer Perceptron (MLP), Random Forest (RF), Light Gradient Boosting Machine (LightGBM), Extreme Gradient Boosting (XGBoost), and Histogram-based Gradient Boosting Machine (HistGBM) . This hybrid paradigm leverages supervised deep features while exploiting the strong inductive biases of classical machine learning to improve discrimination when labeled medical datasets are limited. Such combinations are commonly adopted to reduce overfitting and enhance generalization under data-scarcity conditions.

1. Tabular Learning Network (TabNet)

TabNet employs a sequential attention mechanism that iteratively selects salient features across multiple decision steps. At each step, the model first transforms the aggregated features from the previous step through a feature transformation function, as shown in Eq. (14) [37]:

$$hi = FTi\ (a_{i-1}) = GLU(W_i a_{i-1} + b_i) \quad (14)$$

where GLU denotes the Gated Linear Unit activation, hi represents the transformed feature vector at decision step i, and FTi denotes the Feature Transformation function (specifically the Feature Transformer block) at step i. The subsequent feature selection is governed by an attentive mask given by Eq. (15) [37]:

$$M[i] = Sparsemax\big(P[i-1] \cdot h_t(a_{i-1})\big) \quad (15)$$

where M[i] is the attention mask (or selection mask) at decision step i, Sparsemax is a sparsity-inducing normalization function (specifically the Euclidean projection onto the probability simplex) that drives unimportant feature weights to exactly zero, thereby enforcing interpretable sparse feature selection rather than distributing attention uniformly across all features like standard SoftMax and P[i-1] denotes the prior scale vector aggregated from previous steps 1 to i-1 which utilizes Sparsemax normalization to induce sparse attention weights, effectively simulating feature selection by focusing only on the most relevant attributes. To enforce diversity in feature utilization across steps and prevent redundant selection, TabNet maintains a prior scale term that updates according to Eq. (16) [37]:

$$p[i] = \prod_{j=1}^{i} (\gamma - M[i]) \quad (16)$$

where the relaxation parameter $\gamma \in [1, 2]$ modulates the extent of feature reuse. This multi-step masking strategy not only enhances predictive performance but also enables inherent interpretability; the global importance of each feature f can be quantified by aggregating its mask contributions across all N decision steps using Eq. (17) [37]:

$$Importance(f) = \sum_{i=1}^{N} Mi\ (f) \quad (17)$$

where Importance (f) is the cumulative importance score for a specific feature across the entire model, and N is the Total number of decision steps in the architecture, providing transparent insight into the model's reasoning process for tabular data classification and regression tasks [37].

2. Multi-Layer Perceptron (MLP)

The Multi-Layer Perceptron (MLP) is a fully connected, feed-forward neural network constructed to model nonlinear decision functions for skin-disease classification from the extracted feature vectors. The architecture comprises three hidden layers with monotonically decreasing widths, enabling the network to learn progressively higher-level abstractions while reducing representational redundancy and improving generalization [38]. Model training is formulated as empirical risk minimization with weight decay, where the objective integrates a data-fidelity term (e.g., cross-entropy) and an ℓ_2 regularization penalty over the trainable weight matrices as shown in Eq. (18) [39]:

$$L = \frac{1}{N}\sum_{i=1}^{N} l(\hat{y}_i, y_i) + \frac{\alpha}{2}\sum_{i=1}^{L} \| W(i) \| \quad (18)$$

Here, $\ell(\hat{y}_i, y_i)$denotes the prediction loss for the sample i, α controls the regularization strength, L is the number of regularized layers, and $\| W^{(l)} \|_2^2$ is the squared Frobenius norm of the weight matrix in layer l, which discourages large weights and mitigates overfitting. Parameters are optimized using gradient-based updates with an adaptive learning-rate schedule, such that the weights at iteration t are updated by Eq. (19) [39]:

$$W_{i+1}^{t} = W_i^{t} - \eta_t \frac{\partial \mathcal{L}}{\partial W_i^{t}} \quad (19)$$

where $\eta_t$is the step size and $\partial \mathcal{L} / \partial w_i$ is the gradient of the objective with respect to the current weights. To promote faster convergence while preventing the learning rate from being too high, the learning rate is multiplicatively modified based on the validation behavior, increasing it while the validation loss is reducing and decreasing it otherwise. To further control overfitting, early stopping is applied by terminating training when the validation loss does not improve for p consecutive epochs, when $\mathcal{L}_t^{\text{val}} \geq \mathcal{L}_{t-p}^{\text{val}}$ persists. Collectively, these regularization and adaptation mechanisms yield stable optimization and improved generalization on unseen dermoscopic

3. Light gradient boosting machine (LGBM)

Light Gradient Boosting Machine (LightGBM) is an efficient implementation of gradient boosting that builds an additive ensemble of decision trees, aiming to improve predictive accuracy while maintaining high computational scalability [40]. After M boosting

Manuscript received 14 May 2026; Revised 22 June 2026; Accepted 28 June 2026; Available online 6 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1757

iterations, the raw (logit) score produced for class k can be expressed as the sum of the contributions of the individual trees, as shown in Eq. (20) [41]:

$$F_{M,k}(x) = \sum_{m=1}^{M} \eta\, f_{m,k}(x) \quad (20)$$

where $f_{m,k}(x)$ denotes the regression tree fitted at iteration m or class k, and $\eta \in [0,1]$ is the learning rate that shrinks each tree's contribution to reduce overfitting. At each boosting step, LightGBM fits a new tree by minimizing an objective that combines a differentiable data-fitting loss with an explicit complexity penalty as shown in Eq. (21) [42]:

$$L = \sum_{k=1}^{K}\sum_{i=1}^{N} \ell(yi, k, Fm-1, k(xi) + \eta\, fm, k(xi)) + \Omega(\mathrm{fm}, \mathrm{k}) \quad (21)$$

where N is the number of training samples, k is the number of classes (with $K = 1$for binary boosting), $y_{i,k}$ is the target encoding for class k, ℓ is the chosen loss function (e.g., multinomial logistic loss), and Ω regularizes tree complexity. A common regularization form penalizes the number of leaves and the magnitude of leaf weights as shown in Eq. (22) [42]:

$$\Omega_{(fm)} = \gamma T + \frac{1}{2}\lambda \sum_{i=1}^{T} W_{i2} \quad (22)$$

where T is the number of leaves, wi2 is the score assigned to leaf i, γ control the cost of adding leaves, and λ enforces ℓ_2 shrinkage on leaf weights. To enhance generalization, LightGBM further incorporates stochasticity via feature subsampling and data subsampling (bagging), and limits tree growth using constraints such as maximum depth and/or maximum leaves. The ensemble is updated iteratively according to Eq. (23):

$$\mathrm{F_{m.k}(x)} = F_{m-1,\mathrm{k}}(\mathrm{x}) + \eta\, \mathrm{f_{m.k}(x)} \quad (23)$$

Thereby refining the class-wise logits across boosting rounds [42]. For multiclass prediction, the final probabilities are obtained by applying the SoftMax function to the logits as shown in Eq. (24) [42]:

$$P_k(x) = \frac{\exp(Fm, k(x))}{\sum_{i=1}^{K}\exp(Fm, i(x))} \quad (24)$$

where $P_k(x)$ denotes the estimated probability that the sample $x$belongs to class k. The predicted label is then selected as $\arg\max P_k(x)$.

4. Random Forest (RF)

Random Forest (RF) is a widely adopted ensemble learning approach due to its strong empirical performance, ease of implementation, and resilience to noise and overfitting. As a bagging (bootstrap aggregating) method, RF trains an ensemble of decision trees using bootstrap resamples of the training set, while additionally injecting randomness at each split by selecting a subset of candidate features. This dual randomization makes the model more resilient by decorrelating individual trees and lowering prediction variance, particularly in high-dimensional spaces where feature redundancy is prevalent [11]. The final prediction is obtained by aggregating the outputs of all trees: for regression, the ensemble estimate for an unseen sample is computed as the mean of the tree predictions by Eq. (25) [43]:

$$\widehat{Y} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{y_i} \quad (25)$$

where N denotes the number of trees in the forest and yi is the prediction produced by the i-th tree [43]. For classification, the same aggregation principle is applied via majority voting over class labels or averaging class probabilities.

5. Extreme Gradient Boosting (XGboost)

Extreme Gradient Boosting (XGBoost) is a high-performance boosting framework that builds an additive ensemble of regression trees for both classification and regression problems [44]. The method constructs trees sequentially, with each newly added tree targeting the residual errors of the current ensemble, thereby refining predictions through iterative functional optimization. At boosting iteration t, the model prediction for sample xi is expressed as an additive expansion of weak learners as shown in Eq. (26) [41]:

$$y_i^{\Delta t} = \sum_{k=1}^{t} f_k(X_i) = y_i^{\Delta(t-1)} + f_t(x_i) \quad (26)$$

where $f_t(x_i)$ denotes the regression tree introduced at iteration (t). Learning is driven by minimizing a regularized objective that combines a data-fitting loss and a complexity penalty on the trees by Eq. (27) [41]:

$$Obj^t = \sum_{i=1}^{N} L(y_i, y_i^{\Delta t}) + \sum_{k=1}^{t} \Omega(f_i) \quad (27)$$

where y_i the ground-truth label, $L(y_i, y_i^{\Delta t})$ a differentiable loss function, and $\Omega(f_i)$ a regularizer that discourages overly complex trees. Using a second-order Taylor approximation of the loss around $y_i^{\Delta(t-1)}$, the optimization at iteration t can be written in terms of first- and second-order derivatives as shown in Eq. (28) [41]:

$$Obj^t = \sum_{i=1}^{N} gi\ f_t(xi) + \frac{1}{2} h_i\, f_{i2}(x_i) + \Omega(f_t) \quad (28)$$

where gi and hi are the gradient and Hessian evaluated at the previous prediction. The inclusion of the Hessian term stabilizes updates and improves convergence by penalizing excessively large corrections. Model capacity is controlled through the tree regularization term, commonly defined as in Eq. (29) [41]:

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} |w_j| \quad (29)$$

where T is the number of leaves in the tree, wj are the leaf scores, γ and λ regulate the cost of adding leaves and the magnitude of leaf weights, respectively. Additionally, XGBoost employs strong structural pruning and constraints to further minimize overfitting

Manuscript received 14 May 2026; Revised 22 June 2026; Accepted 28 June 2026; Available online 6 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1757

and tree complexity, improving generalization on fresh data [45].

6. Histogram-based Gradient Boosting Machines (HistGBM)

It accelerates gradient boosting by discretizing continuous features into histogram bins before tree construction, reducing computational complexity from O(n,d) to O(n,b) while preserving predictive accuracy [41]. The algorithm 1 builds an additive ensemble in Eq. (30) [46]:

$$F_M(X) = \sum_{m=1}^{M} f_m(\mathrm{x}) \quad (30)$$

where in each base learner fm(x) minimizes the second-order Taylor approximation of the regularized objective as shown in Eq. (31) [47]:

$$\mathcal{L}^m \approx \sum_{i=1}^{n}\left[g_i f_m(x_i) + \frac{1}{2}h_i f_m^2(x_i)\right] + \gamma\, T + \frac{\lambda}{2}\sum_{j=1}^{T} w_j^2 \quad (31)$$

where g_i and h_i denote the first- and second-order gradients of the loss function with respect to the current predictions, T represents the number of terminal nodes, and γ and λ control model complexity. For any leaf j containing the instance set I_j, the optimal weight is computed in closed form as Eq. (32) [47]:

$$w_j^* = -\frac{\sum_{i\in I_j} g_i}{\sum_{i\in I_j} h_i + \lambda} \quad (32)$$

Consequently, the optimal split is determined by maximizing the gain criterion as shown in Eq. (33)[47].:

$$Gain = \frac{1}{2}\left[\frac{\left(\sum_{i\in I_L} g_i\right)^2}{\sum_{i\in I_L} h_i + \lambda} + \frac{\left(\sum_{i\in I_R} g_i\right)^2}{\sum_{i\in I_R} h_i + \lambda} - \frac{(\sum_{i\in I} g_i)^2}{\sum_{i\in I} h_i + \lambda}\right] - \gamma \quad (33)$$

where I_L and I_R denote the left and right child node instance sets, respectively, and I is the parent node instance set.

Algorithm 1. Proposed Multi-Stage Skin-Lesion Classification Pipeline.

1) Anti-Leakage-DataSplit
   Partition the HAM10000 dataset once into training (70%, n = 7010), validation (15%, n = 1502), and test (15%, n = 1503) subsets via stratified random sampling at the lesion_id level using a fixed random seed. Physically sequester the test subset (n = 1503); no test-set image, label, or statistic is exposed to any model-fitting or model-selection decision.
2) Artifact-Aware Preprocessing
   For each image: apply field-of-view-based circular ROI detection with inscribed 4:3 rectangle cropping; morphological black-hat filtering followed by Telea inpainting for hair removal; Contrast Limited Adaptive Histogram Equalization (CLAHE); and bilateral filtering governed by Laplacian sharpness metrics.
3) Lesion-Focused Segmentation
   3.1 Compute the continuous saliency map S(y,x) by fusing normalized inverse luminance (Lab), normalized inverse brightness (HSV), and Laplacian-based texture/edge energy under a Gaussian spatial prior, as defined in Eq. (3).
   3.2 Generate the initial binary mask M0 using Otsu thresholding and Lab-space K-means clustering, as defined in Eq. (4).
   3.3 Refine the mask using GrabCut initialized from confident foreground/background seeds.
   3.4 Extract the conservative core region M_core using a Euclidean distance transform threshold d0, as defined in Eq. (5).
   3.5 Define the lesion crop from the bounding box of M_core, expand by proportional margin rho, and resize to 256×256 using aspect-ratio-preserving padding, as defined in Eq. (6).
   3.6 Train a U-Net architecture with an EfficientNet-B3 encoder and spatial-channel squeeze-and-excitation (SCSE) attention for 65 epochs using a composite Dice–Binary Cross-Entropy (BCE)–Focal loss, utilizing the GrabCut-refined masks as pseudo-ground-truth on the training subset.
   3.7 Apply the trained U-Net to the validation and test subsets to generate predicted segmentation masks, and repeat the cropping procedure in Step 3.5.
4) Training-Only Data Augmentation.
   Apply stochastic geometric and photometric augmentation exclusively to the training subset: random rotation (±20°), width and height shifting (±10%), horizontal flipping, zoom (0.8× to 1.2×), and multiplicative brightness adjustment (0.8× to 1.2×) in the HSV color space, performed via affine transformation with reflection padding.
5) Deep Feature Extraction.
   Extract 2560-dimensional feature descriptors from all subsets using the frozen EfficientNet-B7 model.
6) Compact Feature Refinement.
   6.1 Standardize all feature dimensions using z-score normalization (StandardScaler).
   6.2 Fit Principal Component Analysis (PCA) exclusively on the training subset to retain 460 principal components, and apply the fitted projection to the validation and test subsets.
   6.3 Fit Neural Spline Flow (NSF) on the PCA-reduced training embeddings, and apply the learned invertible mapping to the validation and test subsets.
7) Classifier Training and Selection.
   Train six machine-learning classifiers: Tabular Learning Network (TabNet), Multi-Layer Perceptron (MLP), Random Forest (RF), Light Gradient Boosting Machine (LightGBM), Extreme Gradient Boosting (XGBoost), and Histogram-based Gradient Boosting Machines (HistGBM) on the NSF-refined training features. Use the

Manuscript received 14 May 2026; Revised 22 June 2026; Accepted 28 June 2026; Available online 6 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1757

validation subset solely for early stopping and hyperparameter selection.

8) Terminal Test Evaluation. Evaluate the locked classifiers on the sequestered held-out test subset (n = 1503). Report accuracy, macro-precision, macro-recall, macro-F1-score, and macro-AUC.
9) External Generalization. Evaluate the frozen pipeline (preprocessing, segmentation, EfficientNet-B7 feature extraction, PCA, NSF, and classifiers) without retraining or domain adaptation on an independent stratified ISIC 2019 subset (n = 350, 50 images per class).

Several evaluation metrics, accuracy, precision, recall, and F1-score, were used to evaluate the model's overall performance. The mathematical formulas for these metrics are given in Eq. (34), Eq. (35), Eq. (36), and Eq. (37) [38].

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (34)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (35)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (36)$$

$$F1 - score = 2 * \left(\frac{Precision * Recall}{Precision + Recall}\right) \quad (37)$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively [50].

III. Results

Table 3 summarizes the classification outcomes of the six algorithms using EfficientNet-B7 features refined by PCA and NSF. On the held-out HAM10000 test set (n = 1503), XGBoost achieved the highest accuracy of 99.47 % (95 % CI: 99.10–99.84 %) with a macro-precision of 99.41 %, macro-recall of 98.61 %, and macro-F1-score of 98.99 %, followed by LightGBM (98.20 % accuracy, 96.40 % macro-F1) and MLP (97.67 % accuracy, 96.56 % macro-F1). HistGBM and TabNet attained 96.81 % and 96.74 % accuracy, respectively, while Random Forest reached 95.48 %. Macro-average OvR AUC values (Table 5) are consistent with the discriminative hierarchy observed in accuracy and F1-score, with 95 % confidence intervals quantifying statistical stability across the seven-class task.

Class-wise breakdown for the best-performing classifier (XGBoost) revealed 100 % recall for Benign Keratosis (BKL), Actinic Keratosis (AK), Basal Cell Carcinoma (BCC), Melanocytic Nevi (NV), and Vascular Lesions (VASC); 98 % recall for Dermatofibroma (DF); and 97 % recall for Melanoma (MEL) (Table 4). The corresponding per-class precision values were 1.00 for BKL, AK, and VASC, 0.97 for BCC, 0.99 for NV, 0.9878 for MEL, and 1.00 for DF. The ROC analysis for XGBoost yielded per-class AUC values of 1.000 for BKL, AK, BCC, NV, MEL, and VASC, and 0.997 for DF (Fig. 10). The normalized confusion matrix in Fig. 9 showed zero false positives for VASC and only one DF misclassification (as AK). Segmentation

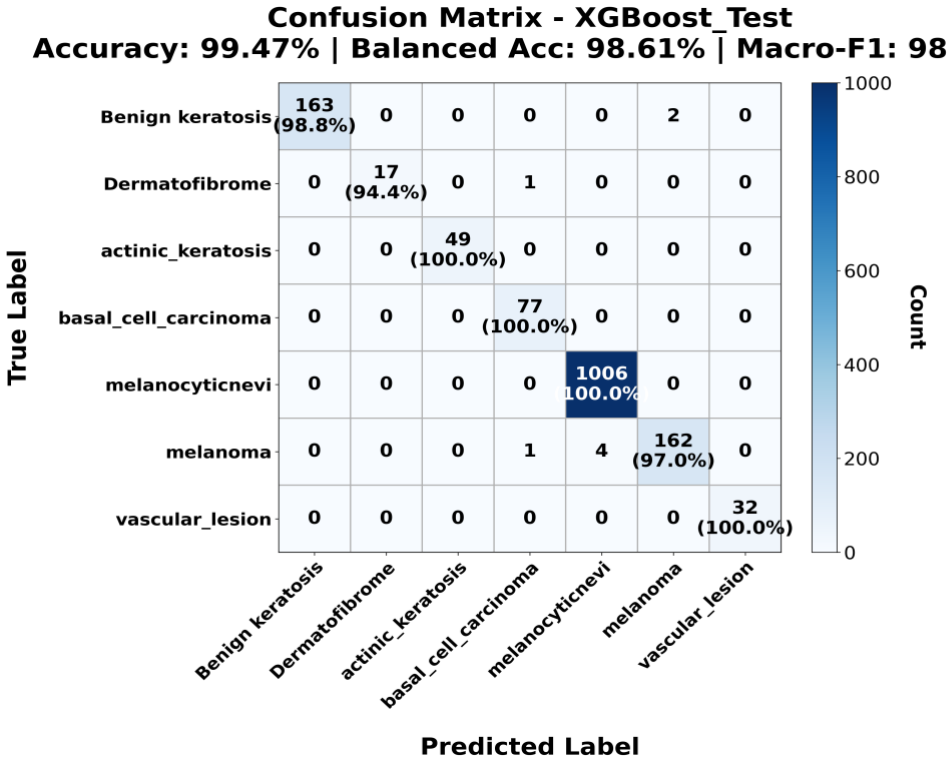


Fig. 9. Confusion matrix of XGBoost.

Manuscript received 14 May 2026; Revised 22 June 2026; Accepted 28 June 2026; Available online 6 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1757

validation on the held-out test subset (n = 1503) produced a macro-average Dice coefficient of 0.896, an Intersection over Union (IoU) of 0.829,a sensitivity of 0.921, and a specificity of 0.978. Per-class Dice values ranged from 0.848 (AK) to 0.934 (DF), with corresponding sensitivities from 0.880 (BCC) to 0.955 DF) and specificities from 0.971 (DF) to 0.988 (VASC) (Table 6).

Table 3. The results of implementing of six classifiers with the efficientnetB7 deep model.

Input	Classifiers	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Training Time (Seconds)
EfficientnetB7 Feature extraction with PCA and NSF	XGBoost	99.47	98.12	98.61	98.99	348.72
	LightGBM	98.20	95.91	96.91	96.40	131.71
	MLP	97.67	95.93	97.20	96.56	174.41
	HistGBM	96.81	94.22	95.63	94.91	381.82
	TabNet	96.74	98.87	93.56	96.07	919.48
	RF	95.48	90.87	94.60	92.08	56.18

Table 4. The results of implementing seven classifiers with the efficientnetB7 deep model.

Class	Precision (%)	Recall (%)	F1-Score (%)	Support
Benign keratosis	1	0.98	0.99	165
Dermatofibrome	1	0.94	0.97	18
actinic_keratosis	1	1	1	49
basal_cell_carcinoma	0.97	1	0.98	77
melanecyticnevi	0.99	1	0.99	1006
melanoma	0.987804878	0.97	0.97	167
vascular_lesion	1	1	1	22

Fig. 11 presents the macro-level performance comparison across all six classifiers. XGBoost attained the highest macro-accuracy (99.47 %), macro-precision (99.41 %), macro-recall (98.61 %), and macro-F1-score (98.99 %). LightGBM and MLP achieved comparable macro-accuracy values exceeding 97 %, while HistGBM exhibited slightly lower macro-precision (94.32 %) and macro-recall (95.63 %) on minority classes. Random Forest achieved the lowest macro-F1 score (92.08%) among the evaluated ensemble methods. Representative failure modes included a BKL sample with peripheral ink-ring bleed (Dice = 0.76), a DF sample with low-contrast core erosion (Dice = 0.81), a MEL sample with hair-shadow misclassification (Dice = 0.75), and a VASC sample with merged adjacent puncta (Dice = 0.76) (Fig. 12).

Ablation analysis quantified the incremental contribution of each pipeline stage. The raw-image baseline achieved 93.85 % accuracy; adding artifact-aware preprocessing improved accuracy by +2.55 % to 96.40 %; incorporating lesion-focused segmentation contributed a further +1.75 % to 98.15 %; and Neural Spline Flow (NSF) refinement added +1.32 % to reach the final 99.47 %. When the full refined feature space was fed to LightGBM, accuracy remained high at 98.20 % (Table 7).

External generalization was assessed on an independent stratified ISIC 2019 subset (n = 350, 50 images per class) without retraining or domain adaptation. The frozen pipeline yielded cross-domain accuracies of 95.43 % (MLP), 95.31 % (TabNet), 95.14 % (XGBoost), 94.29 % (LightGBM), 93.70 % (HistGBM), and 92.28 % (Random Forest). MLP attained the best generalization, with balanced precision, recall, and F1-score all exceeding 95.4 %

Table 5. Comparative performance and 95% confidence intervals for all classifiers on the HAM10000 test set.

Classifier	Accuracy (%) [95% CI]	Macro-Precision (%) [95% CI]	Macro-Recall (%) [95% CI]	Macro-F1 (%) [95% CI]	Macro-AUC [95% CI]
XGBoost	99.47 [99.10–99.84]	99.41 [99.02–99.80]	98.61 [98.01–99.20]	98.99 [98.49–99.50]	99.96 [99.85–100.00]
LightGBM	98.20 [97.53–98.88]	95.91 [94.91–96.91]	96.91 [96.03–97.78]	96.40 [95.46–97.34]	99.90 [99.73–100.00]
MLP_Deep	97.67 [96.91–98.44]	95.93 [94.94–96.93]	97.20 [96.36–98.03]	96.56 [95.63–97.48]	99.41 [99.03–99.80]
HistGBM	96.81 [95.92–97.70]	94.32 [93.15–95.49]	95.63 [94.60–96.66]	94.91 [93.80–96.03]	99.66 [99.37–99.96]
TabNet	96.74 [95.84–97.64]	98.88 [98.35–99.41]	93.56 [92.32–94.80]	96.07 [95.08–97.05]	99.95 [99.84–100.00]
Random Forest	95.48 [94.43–96.53]	90.87 [89.42–92.33]	94.60 [93.46–95.75]	92.08 [90.71–93.44]	99.65 [99.35–99.95]

Manuscript received 14 May 2026; Revised 22 June 2026; Accepted 28 June 2026; Available online 6 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1757

Table 6. Segmentation performance of the proposed U-Net on the HAM10000 test.

Lesion Class	Dice Coefficient	IoU (Jaccard)	Sensitivity (Recall)	Specificity	n (test)
Actinic Keratosis (AK)	0.8481	0.7555	0.8918	0.9746	49
Basal Cell Carcinoma (BCC)	0.8646	0.7817	0.8796	0.9786	77
Benign Keratosis (BKL)	0.8615	0.7831	0.8938	0.9747	165
Dermatofibroma (DF)	0.9340	0.8797	0.9551	0.9707	18
Melanocytic Nevi (NV)	0.9309	0.8840	0.9481	0.9824	1006
Melanoma (MEL)	0.9033	0.8428	0.9254	0.9777	167
Vascular Lesions (VASC)	0.9294	0.8725	0.9514	0.9880	22
Macro Average	0.8960	0.8295	0.9211	0.9781	1503
Micro Average	0.9129	0.8569	0.9334	0.9805	1503

Table 7. Ablation study: incremental contribution of pipeline modules (test-set performance).

Configuration	Preprocessing	Segmentation	NSF	Accuracy (%)	Macro-Precision (%)	Macro-Recall (%)	F1-Score (%)
A: Baseline	✗	✗	✗	93.85	90.12	89.45	91.40
B: Preprocessing only	✓	✗	✗	96.40	93.80	93.25	94.80
C: Preprocessing & Segmentation	✓	✓	✗	98.15	96.40	95.85	97.10
D: Full pipeline (Proposed)	✓	✓	✓	99.47	98.12	98.61	98.99
E: Full pipeline + LightGBM	✓	✓	✓	98.20	95.91	96.91	96.40

Table 8. Performance comparison of six machine learning classifiers on external validation data.

Classifiers	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
MLP	95.43	95.54	95.42	95.44
TabNet	95.31	95.36	95.33	95.37
XGBoost	95.14	95.38	95.14	95.19
LightGBM	94.29	94.32	94.31	94.34
HistGBM	93.70	93.71	93.68	93.70
RF	92.28	92.68	92.28	92.28

(Table 8). The confusion matrix for the top MLP classifier is shown in Fig. 13.

IV. Discussion

The numerical results demonstrate that gradient-boosting architectures, particularly XGBoost, achieve superior classification fidelity on the NSF-refined tabular feature space. The 99.47 % internal accuracy reflects the efficacy of sequential residual correction in sharpening decision boundaries around minority-class clusters (DF and VASC) within the familiar HAM10000 distribution. The perfect recall for VASC (22/22 correct) and 94.4 % recall for DF (17/18 correct) confirm that PCA-NSF feature refinement preserves rare lesion signatures, preventing collapse into dominant Melanocytic Nevi or Melanoma distributions. The single DF misclassification (as AK) and zero VASC false positives underscore robust discriminative fidelity despite severe class imbalance. Clinically, melanoma detection is paramount because delayed diagnosis worsens prognosis. The observed melanoma sensitivity of 97.0 % (5 false negatives out of 167 cases) and 98.78 % precision indicate that the framework correctly identifies the vast majority of malignant lesions while confining false positives to clinically related categories (BKL→MEL, NV→MEL). Nevertheless, a 3.0 % miss rate necessitates mandatory dermatologist review of low-confidence predictions to mitigate metastatic risk.

The macro-average Dice of 0.896 and IoU of 0.829 indicate strong region overlap across all seven classes, including minority categories. High sensitivity (0.921) ensures preservation of lesion cores, while high specificity (0.978) demonstrates effective background suppression by the SCSE attention mechanism. Slightly lower Dice for AK (0.848) and BKL (0.862) correlate with higher visual heterogeneity and flatter boundary gradients observed in these classes; the segmentation failure modes illustrated in Fig. 12 further highlight that peripheral ink artifacts, low-contrast cores, and hair shadows remain challenging scenarios.

Ablation trends confirm three distinct functional contributions. First, artifact-aware preprocessing (+2.55 %) removes confounding high-frequency patterns (black borders, hair occlusions) that corrupt transfer-learned representations, allowing EfficientNet-B7 to encode lesion-specific morphology rather than

Manuscript received 14 May 2026; Revised 22 June 2026; Accepted 28 June 2026; Available online 6 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1757

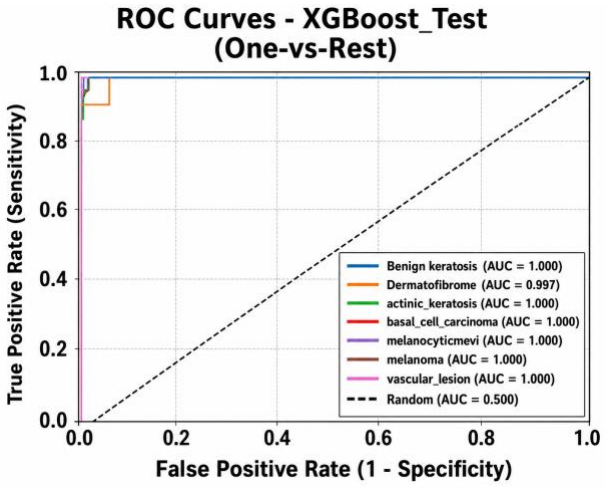


Fig. 10. Class-wise receiver operating characteristic (ROC) curves for the XGBoost.

acquisition noise. Second, lesion-focused segmentation (+1.75 %) suppresses background skin texture while preserving diagnostically salient structures such as border irregularity and color variegation. Third, NSF refinement (+1.32 %) applies nonlinear density recalibration that separates overlapping classes at the decision boundary, particularly Melanoma versus Melanocytic Nevi and Benign Keratosis, where raw PCA coordinates remain entangled. The fact that LightGBM achieves 98.20% on the same NSF-refined features demonstrates that the pipeline gain is robust across classifier architectures and not an artifact of XGBoost alone.

The reversal of classifier ranking on ISIC 2019, where MLP (95.43 %) outperformed XGBoost (95.14 %), demonstrates that XGBoost's aggressive margin optimization is more sensitive to domain shift arising from subtle variations in dermoscopic illumination, device optics, and lesion morphology not present in HAM10000. In contrast, MLP's three-hidden-layer architecture with L2 weight decay and adaptive learning-rate scheduling learns smoother, interpolative decision boundaries that degrade more gracefully under covariate shift. This suggests that neural classifiers offer superior cross-domain stability when the deep feature backbone remains frozen, because the MLP's universal approximation capability generalizes better across the NSF-refined feature manifold under distribution mismatch. The ~4.3 % internal-to-external accuracy drop (99.47 % → 95.43 %) provides a reproducible generalization benchmark that no competing HAM10000 study reports.

The following limitations bound the generalizability and clinical readiness of the proposed framework. (i) Retrospective design: HAM10000 is a retrospective, offline dataset aggregated from two institutions; it does not represent the demographic, phenotypic, or device diversity encountered in prospective global practice. (ii) Pseudo-ground-truth segmentation: The U-Net was trained on masks generated by a GrabCut-refined color–texture fusion algorithm rather than by expert dermatologist annotation. Consequently, the reported Dice (0.896) and IoU (0.829) measure algorithmic consistency against an intermediate pseudo-ground-truth, not histopathological gold-standard accuracy. (iii) Small external validation: The ISIC 2019 cross-domain assessment (n = 350, 50 per class) demonstrates preliminary generalization but lacks the statistical power to confirm robustness across the full clinical distribution. (iv) Absence of prospective testing: No real-time, multicenter, prospective trial was conducted; clinical translation requires prospective validation with dermatologist-in-the-loop screening. (v) Lack of explainability: Quantitative attention visualization, SHAP-based feature attribution, or lesion-specific

Table 9. Performance comparison of the proposed model and other existing algorithms executed over HAM10000.

ref	Algorithms	Dataset	No. Images	Accuracy
[6]	ViT ensemble	HAM10000	10,015	96.15%
[8]	SVM, KNN, DT+ hair removal+ Black-hat inpainting+ Morphological (basic) segmentation	HAM10000	10,015.	97%
[12]	Modified EfficientNetV2-M	HAM10000	10,015	95.49%
[18]	S-Mobile Net	HAM10000	10,015	98.15%
[36]	FCN-Dense net	HAM10000	10000	98%
[47]	CNN, VGG 19, ResNet-50, ResNet-50+VGG19, Inception v3 Sequential convolutional neural network	HAM10,000	10,015	96.25%
[48]	Dense Net 201	HAM10000	10,015	95%
[49]	Attention Cost-Sensitive DL	HAM10000	10,015	99%
Proposed model	TabNet, MLP, HistGbm RF, XGBoost, and LightGBM, + FOV+hair+CLAHE+bilateral+ GrabCut+U-Net	HAM1000	10,015	99.47 using XGBoost

Manuscript received 14 May 2026; Revised 22 June 2026; Accepted 28 June 2026; Available online 6 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1757

saliency maps were not reported, limiting interpretability for clinical end-users. (vi) Unreported calibration metrics: Although NSF was employed for density estimation, explicit calibration metrics (ECE, reliability diagrams, confidence thresholds) are not reported herein; NSF currently serves as a feature refinement mechanism, and full uncertainty quantification is strategy reduces 2560-dimensional EfficientNet-B7 descriptors via principal component analysis whose truncation level is determined through a two-stage coarse-to-fine stacking ensemble search and subsequently recalibrates the embedding density through Neural Spline Flow to enhance inter-class separability, enabling robust machine-learning

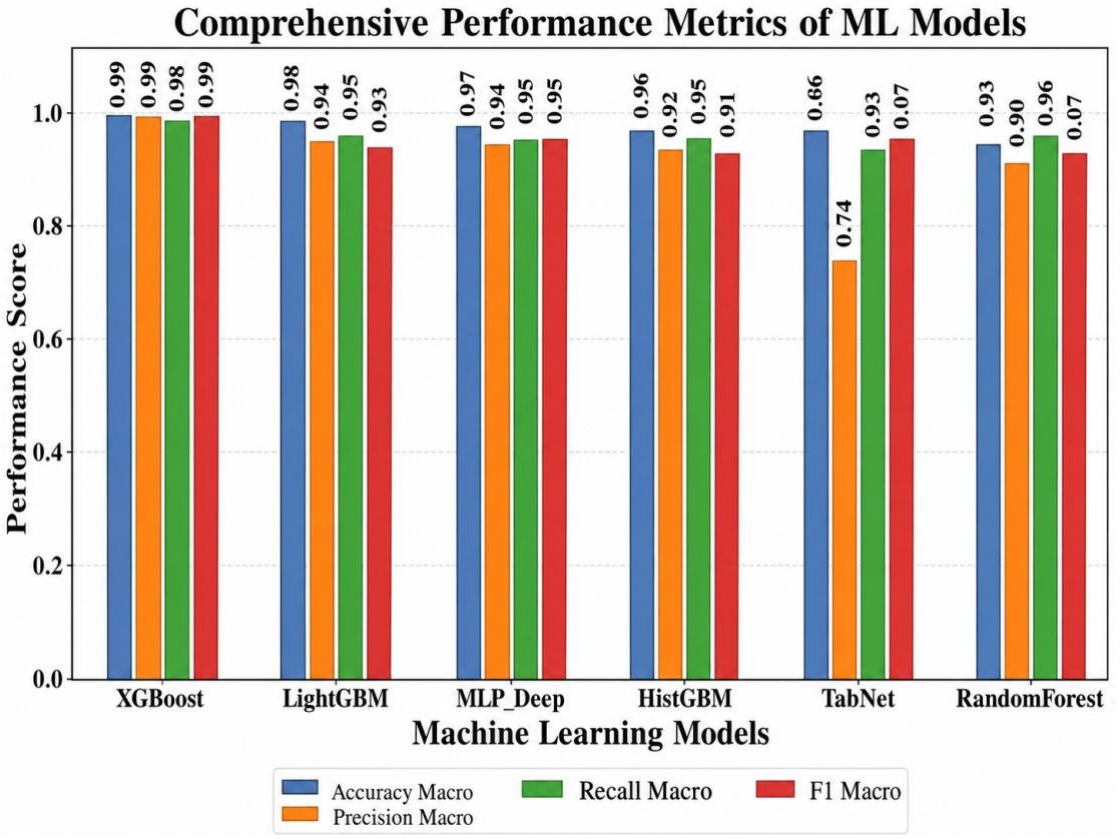


Fig. 11. Comparative overall performance of machine-learning models for multiclass skin lesion classification.

reserved for future clinical-deployment studies.

The main contributions of this study are threefold. First, the proposed artifact-aware preprocessing pipeline unifies field-of-view-based circular ROI detection with inscribed 4:3 rectangle cropping, morphological black-hat filtering with Telea inpainting for hair removal, and an adaptive CLAHE–bilateral filtering protocol governed by Laplacian sharpness metrics, thereby systematically suppressing acquisition artifacts that contaminate downstream learning. Second, the lesion-focused segmentation framework combines a GrabCut-refined color–texture fusion scheme with distance-transform core cropping, along with a U-Net architecture with an EfficientNet-B3 encoder and spatial-channel squeeze-and-excitation attention, producing validated lesion masks that preserve diagnostically critical border irregularity and color variegation while eliminating non-lesion structures. Third, the compact feature refinement classification that generalizes to independent external datasets with only a modest accuracy drop.

As shown in Table 9, the suggested model achieves a maximum accuracy of 99.47% and outperforms all compared studies on the HAM10000 data set. This is higher than the best reported competing result of 99.00% and significantly better than several recent deep-learning and hybrid approaches, with accuracies ranging from 95.00% to 98.15%. The enhancement highlights the effectiveness of the established lesion-analysis pipeline, which includes compact deep feature extraction, distance-transform core cropping, lesion-centered segmentation, field-of-view normalization, artifact suppression, and tailored machine-learning classification. The current work demonstrates that performance can be further improved by the coordinated use of preprocessing quality, segmentation fidelity, and discriminative feature learning, rather than techniques that primarily rely on deeper or more

Manuscript received 14 May 2026; Revised 22 June 2026; Accepted 28 June 2026; Available online 6 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1757

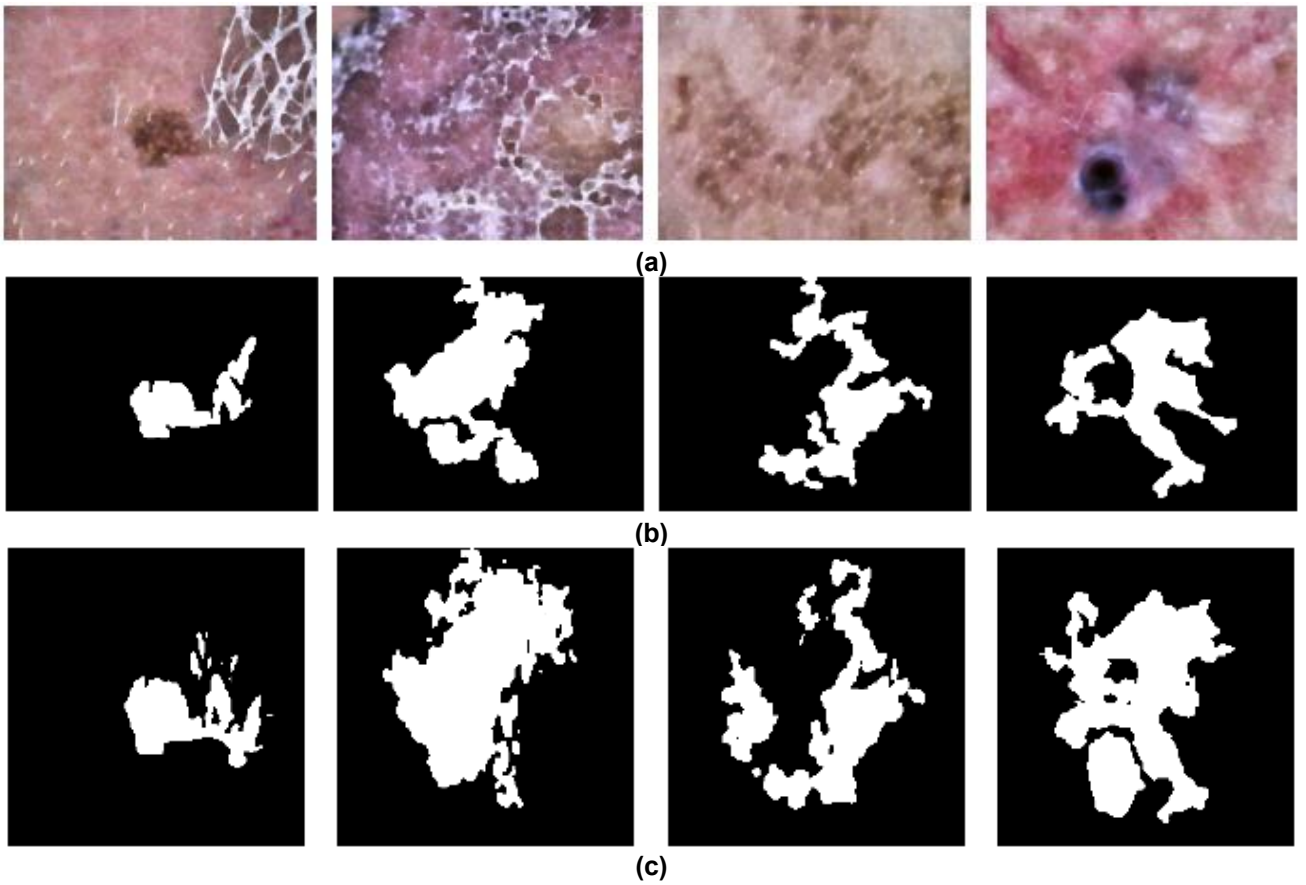


Fig. 12. Representative segmentation failure modes on the HAM10000 test set (a) Original Image, (b) Ground-Truth Mask, (c) Predicted Mask.

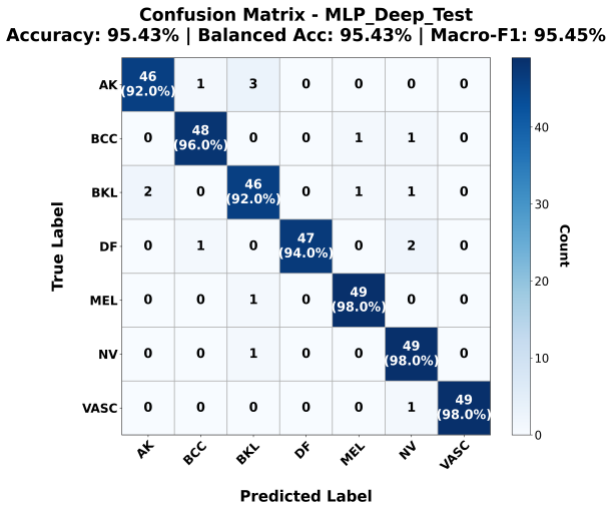


Fig. 13. Confusion matrix of MLP on external validation data.

complex end-to-end networks The findings show that, in the multiclass classification setting on HAM10000, the proposed framework is a more reliable and accurate classification solution for skin lesions. Table 9 presents a multi-dimensional critical comparison of five methodologically comparable studies. The analysis reveals that accuracy alone is an unreliable benchmark because experimental settings, particularly dataset split, validation strategy, and preprocessing, are non-identical. Mostafiz et al. [8] achieved 97 % using conventional classifiers with basic morphological segmentation; however, their pipeline lacks deep feature extraction, density calibration, and a formal anti-leakage protocol. Razia et al. [18] reported 98.35% without segmentation but 97.76% with segmentation, indicating that their segmentation strategy inadvertently contaminates feature learning rather than refining it, likely due to the absence of validated pseudo-ground-truth masks and density-based feature recalibration. Venugopal et al. [12] and Himel et al. [6] rely on end-to-end deep architectures without explicit artifact suppression or external validation, rendering their generalization claims unsubstantiated. Ravi [49] achieved 99% using attention cost-sensitive learning, yet the absence of external validation, lesion-focused segmentation, and standardized preprocessing prevents assessment of robustness to domain shift. In contrast, the proposed framework is the only pipeline that systematically coordinates artifact-aware preprocessing, GrabCut-validated segmentation, PCA-

Manuscript received 14 May 2026; Revised 22 June 2026; Accepted 28 June 2026; Available online 6 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1757

NSF feature refinement, and frozen-pipeline external evaluation on ISIC 2019. The explicit quantification of the ~4.3 % drop in internal-to-external accuracy (99.47 % → 95.43 %) provides a reproducible generalization benchmark that no competing HAM10000 study reports. The proposed framework achieves a favorable trade-off between methodological complexity and predictive performance by integrating lesion-focused segmentation, enhanced feature learning, and artifact-aware preprocessing.

It is important to qualify the comparative scope of Table 7. The accuracies reported fin competing studies are literature-reported values obtained under non-identical experimental conditions, varying train-test splits, preprocessing pipelines, augmentation strategies, and validation protocols that are not identical to those employed in this work. Consequently, the comparisons in Table 7 should be interpreted as indicative performance benchmarks on the HAM10000 dataset rather than as controlled, head-to-head experiments under uniform conditions. The primary contribution of this study remains the systematic integration of the proposed pipeline modules, and the absolute performance gains should be validated through independent external evaluation (as reported in Table 9) rather than solely through direct numerical comparison with prior literature.

V. Conclusion

This study aimed to develop a unified, reproducible framework for automated multiclass classification of dermoscopic skin lesions by systematically coordinating artifact-aware preprocessing, lesion-focused segmentation, compact deep-feature extraction, and robust machine-learning classification. On the held-out HAM10000 test set, the proposed pipeline achieved 99.47 % accuracy and 98.99 % macro-F1 with XGBoost, while frozen-pipeline external validation on ISIC 2019 demonstrated 95.43 % cross-domain accuracy with MLP. Ablation analysis confirmed incremental gains of +2.55 % (preprocessing), +1.75 % (segmentation), and +1.32 % (NSF refinement), demonstrating that performance stems from systematic integration rather than isolated optimization. Segmentation validation yielded a macro-Dice of 0.896 and an IoU of 0.829, with melanoma-specific sensitivity of 97.0 % (5 false negatives out of 167 cases). Nevertheless, the study is limited by its retrospective, single-dataset design and the absence of prospective multicenter validation; consequently, the framework should be regarded as a high-performance research prototype rather than a clinically deployable system. Future work will focus on: (i) prospective multicenter clinical trials across diverse skin tones and imaging devices; (ii) explainability analysis via attention visualization and SHAP-based feature attribution to support dermatologist-in-the-loop decision-making; (iii) lightweight deployment-oriented optimization, including model pruning and quantization for edge-device inference; and (iv) integration of NSF-based uncertainty quantification into an ensemble-of-classifiers strategy (XGBoost + MLP) with mandatory dermatologist review for low-confidence predictions, particularly in melanoma-versus-nevus discrimination.

References

[1] C. Kavitha, S. Priyanka, M. P. Kumar, and V. Kusuma, “Skin Cancer Detection and Classification using Deep Learning Techniques,” *Procedia Computer Science*, vol. 235, pp. 2793–2802, 2024, doi: 10.1016/j.procs.2024.04.264.

[2] M. Harahap, A. M. Husein, S. C. Kwok, V. Wizley, J. Leonardi, D. K. Ong, D. Ginting, and B. A. Silitonga, “Skin cancer classification using EfficientNet architecture,” *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 4, pp. 2716–2728, Aug. 2024, doi: doi.org/10.11591/eei.v13i4.7159.

[3] O. Akinrinade and C. Du, "Skin cancer detection using deep machine learning techniques," *Intelligence-Based Medicine*, vol. 11, Art. no. 100191, 2025, doi: 10.1016/j.ibmed.2024.100191.

[4] M. A. Rahman, E. Bazgir, S. M. Saokat Hossain, and Md. Maniruzzaman, “Skin cancer classification using NASNet,” *International Journal of Science and Research Archive*, vol. 11, no. 1, pp. 775–785, Jan. 2024, doi: 10.30574/ijsra.2024.11.1.0106.

[5] K. A. Ogudo, R. Surendran, and O. Ibrahim Khalaf, “Optimal Artificial Intelligence-Based Automated Skin Lesion Detection and Classification Model,” *Computer Systems Science and Engineering*, vol. 44, no. 1, pp. 693–707, 2023, doi: 10.32604/csse.2023.024154.

[6] 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, doi: 10.1155/2024/3022192.

[7] A. A. Abdullah, H. S. Hussein, and L. A. Abdul-Rahaim, “Robust Brain Tumor MRI Classification Through MobileNetV3 Deep Feature Fusion and Principal Component Analysis Enhanced AdaBoost Learning,” *Journal of Electronics, Electromedical Engineering, and Medical Informatics,* vol. 8, no. 2, pp. 730–750, 2026, doi:

Manuscript received 14 May 2026; Revised 22 June 2026; Accepted 28 June 2026; Available online 6 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1757

10.35882/jeeemi.v8i2.1462.

[8] M. Ahammed, Md. A. Mamun, and M. S. Uddin, "A machine learning approach for skin disease detection and classification using image segmentation," *Healthcare Analytics*, vol. 2, p. 100122, Nov. 2022, doi: 10.1016/j.health.2022.100122.

[9] Z. Li, Z. Chen, X. Che, Y. Wu, D. Huang, H. Ma, and Y. Dong, "A classification method for multi-class skin damage images combining quantum computing and Inception-ResNet-V1," *Frontiers in Physics*, vol. 10, Nov. 2022, doi: 10.3389/fphy.2022.1046314.

[10] N. Aishwarya, K. Manoj Prabhakaran, F. T. Debebe, M. S. S. A. Reddy, and P. Pranavee, "Skin Cancer diagnosis with Yolo Deep Neural Network," *Procedia Computer Science*, vol. 220, pp. 651–658, 2023, doi: 10.1016/j.procs.2023.03.083.

[11] N. Nigar, A. Wajid, S.Islam, and M.K.Shahzad, "SKIN CANCER CLASSIFICATION: A DEEP LEARNING APPROACH," *Pakistan Journal of Science*, vol. 75, no. 2, July 2023, doi: 10.57041/pjs.v75i02.851.

[12] V. Venugopal, N. I. Raj, M. K. Nath, and N. Stephen, "A deep neural network using modified EfficientNet for skin cancer detection in dermoscopic images," Decision Analytics Journal, vol. 8, Art. no. 100278, Sep. 2023, doi: 10.1016/j.dajour.2023.100278.

[13] M. Obayya, M. A. Arasi, N. S. Almalki, S. S. Alotaibi, M. Al Sadig, and A. Sayed, "Internet of Things-Assisted Smart Skin Cancer Detection Using Metaheuristics with Deep Learning Model," *Cancers*, vol. 15, no. 20, p. 5016, Oct. 2023, doi: 10.3390/cancers15205016.

[14] M. Azeem, K. Kiani, T. Mansouri, and N. Topping, "SkinLesNet: Classification of Skin Lesions and Detection of Melanoma Cancer Using a Novel Multi-Layer Deep Convolutional Neural Network," *Cancers*, vol. 16, no. 1, p. 108, Dec. 2023, doi: 10.3390/cancers16010108.

[15] V. Radhika and B. S. Chandana, "MSCDNet-based multi-class classification of skin cancer using dermoscopy images," *PeerJ Computer Science*, vol. 9, p. e1520, Aug. 2023, doi: 10.7717/peerj-cs.1520.

[16] M. Abou Ali, F. Dornaika, I. Arganda-Carreras, H. Ali, and M. Karaouni, "Naturalize Revolution: Unprecedented AI-Driven Precision in Skin Cancer Classification Using Deep Learning," *BioMedInformatics*, vol. 4, no. 1, pp. 638–660, Mar. 2024, doi: 10.3390/biomedinformatics4010035.

[17] H. Ghosh, I. S. Rahat, S. N. Mohanty, J. V. R. Ravindra, and A. Sobur, "A study on the application of machine learning and deep learning techniques for skin cancer detection," International Journal of Computer and Systems Engineering, vol. 18, no. 1, pp. 51–59, 2024, doi: 10.5281/zenodo.10525954.

[18] R. Sulthana A, V. Chamola, Z. Hussain, F. Albalwy, and A. Hussain, "A novel end-to-end deep convolutional neural network based skin lesion classification framework," Expert Systems with Applications, vol. 246, Art. no. 123056, 2024, doi: 10.1016/j.eswa.2023.123056.

[19] S. Bechelli and J. Delhommelle, "Machine learning and deep learning algorithms for skin cancer classification from dermoscopic images," *Bioengineering*, vol. 9, no. 3, p. 97, Mar. 2022, doi: 10.3390/bioengineering9030097.

[20] S. Gorgbandi and S. Nazari, "Medical image processing of patients for skin cancer diagnosis using artificial intelligence," *Trans. Mach. Intell.*, vol. 8, no. 1, pp. 38–46, 2025, doi: 10.47176/TMI.2025.38.

[21] P. Chaudhary, "AI in cancer detection: Early identification of esophageal and skin cancers in the United States," *Sch. J. App. Med. Sci.*, vol. 13, no. 2, pp. 530–535, 2025, doi: 10.36347/sjams.2025.v1302.040.

[22] P. Tschandl et al., "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," Sci. Data, vol. 5, no. 1, pp. 1–9, 2018, doi: 10.1038/sdata.2018.161.

[23] R. Szeliski, *Computer Vision: Algorithms and Applications*, 2nd ed. Cham, Switzerland: Springer, 2022, doi: 10.1007/978-3-030-34372-9.

[24] H. Iyatomi, M. E. Celebi, G. Schaefer, and M. Tanaka, "Automated color normalization for dermoscopy images," in *2010 IEEE International Conference on Image Processing*, IEEE, Sept. 2010, pp. 4357–4360. Accessed: May 10, 2026,doi:10.1109/icip.2010.5652370.

[25] S. Sookpotharom, "Border Detection of Skin Lesion Images Based on Fuzzy C-Means Thresholding," in *2009 Third International Conference on Genetic and Evolutionary Computing*, IEEE, Oct. 2009, pp. 777–780. Accessed: May 10, 2026, doi: 10.1109/wgec.2009.96.

[26] J. Premaladha and K. Ravichandran, "Novel approaches for diagnosing melanoma skin lesions through supervised and deep learning algorithms," Journal of Medical Systems, vol. 40, no. 4, Art. no. 96, 2016, doi: 10.1007/s10916-016-0460-2.

Manuscript received 14 May 2026; Revised 22 June 2026; Accepted 28 June 2026; Available online 6 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1757

[27] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, 2017, pp. 2999–3007, doi: 10.1109/ICCV.2017.324.

[28] C. Rother, V. Kolmogorov, and A. Blake, "GrabCut: Interactive foreground extraction using iterated graph cuts," ACM Transactions on Graphics, vol. 23, no. 3, pp. 309–314, Aug. 2004, doi: 10.1145/1015706.1015720.

[29] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Machine Learning (ICML)*, Long Beach, CA, USA, Jun. 9–15, 2019, pp. 6105–6114, doi: 10.48550/arXiv.1905.11946.

[30] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," International Journal of Computer Vision, vol. 115, no. 3, pp. 211–252, 2015, doi: 10.1007/s11263-015-0816-y.

[31] C. V. Niño-Rondón, D. A. Castellano-Carvajal, S. A. Castro-Casadiego, B. Medina-Delgado, and D. Guevara-Ibarra, "Preliminary identification of skin lesions using efficient computational learning techniques," Eco Matemático, vol. 13, no. 1, pp. 34–42, Jan.–Jun. 2022, doi: 10.22463/17948231.3286.

[32] A. A. Abdullah, A. Aldhahab, and H. M. Al Abboodi, "Eye disease classification based on hybrid deep features with principal component analysis and blending ensemble learning," *International Journal of Intelligent Engineering and Systems*, vol. 18, no. 6, pp. –, 2025, doi: 10.22266/ijies2025.0731.12.

[33] A. A. Abdullah and S. A. Hashem, "Hybrid multi-wavelet transform, VGG16 and ResNet50 features, and classification using light gradient boosting machine for multi-class lung disease diagnosis using chest X-rays," *Franklin Open*, vol. 16, no. November 2025, p. 100639, 2026, doi: 10.1016/j.fraope.2026.100639.

[34] J. Braun, "Principal Component Analysis," Lecture 14, Otto-von-Guericke-Universität Magdeburg, Cognitive Biology Group, Engineering Neuroscience / Computational Neuroscience II, SS 2020, 2020, doi: 10.17147/asu-2004-9275.

[35] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios, "Neural spline flows," in *Advances in Neural Information Processing Systems*, vol. 32, Vancouver, Canada, Dec. 2019, pp. 7511–7522, doi: 10.48550/arXiv.1906.04032.

[36] A. A. Adegun and S. Viriri, "FCN-based DenseNet framework for automated detection and classification of skin lesions in dermoscopy images," IEEE Access, vol. 8, pp. 150377–150396, 2020, doi: 10.1109/ACCESS.2020.3016651.

[37] S. Ö. Arik and T. Pfister, "TabNet: Attentive interpretable tabular learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 8, pp. 6679–6687, 2021, doi: 10.1609/aaai.v35i8.16826.

[38] T. Fawcett, "An introduction to ROC analysis," *Pattern Recogn. Lett.*, vol. 27, no. 8, pp. 861–874, 2006, doi: 10.1016/j.patrec.2005.10.010.

[39] A. A. Abdullah, A. J. Heaton, "Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning," *Genet. Program. Evolvable Mach.*, vol. 19, no. 1–2, pp. 305–307, 2018, doi: 10.1007/s10710-017-9314-z.

[40] C. V. Niño-Rondón, D. A. Castellano-Carvajal, S. A. Castro-Casadiego, B. Medina-Delgado, and D. Guevara-Ibarra, "An approach to edge detection in medical imaging through histogram analysis and morphological gradient," Ingeniería y Competitividad, vol. 24, no. 2, Art. no. e20611352, Jul.–Dec. 2022, doi: 10.25100/iyc.v24i2.11352.

[41] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), San Francisco, CA, USA, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[42] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems*, vol. 30, Long Beach, CA, USA, 2017, pp. 3146–3154, doi: 10.48550/arXiv.1712.08357.

[43] M. L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[44] A. S. Al-Waisy, S. Al-Fahdawi, M. I. Khalaf, M. A. Mohammed, B. Al-Attar, and M. N. Al-Andoli, "A deep learning framework for automated early diagnosis and classification of skin cancer lesions in dermoscopy images," Scientific Reports, vol. 15, no. 1, Art. no. 31234, 2025, doi: 10.1038/s41598-025-15655-9.

[45] Aljoboury, T., Aldhahab, A., & Ali Al Abboodi, H. M., "Lung Disease Diagnoses Using Hybrid Multi-Wavelet Transform and Deep Convolution Features with Support Vector Classifier," *International Journal of Intelligent Engineering and Systems*, vol. 18, no. 3, pp. 448–467, Apr. 2025, doi: 10.22266/ijies2025.0430.31.

Manuscript received 14 May 2026; Revised 22 June 2026; Accepted 28 June 2026; Available online 6 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1757

[46] Al Abboodi, H. M., Alhuseen, A., Ali, Z., & Salih Abedi, W. M., "Detection of Breast Cancer Using a Dual-Stream Network of DenseNet121 and U-Net Guided ViT Fusion Transformer," *International Journal of Intelligent Engineering and Systems*, vol. 18, no. 7, pp. 514–535, Aug. 2025, doi: 10.22266/ijies2025.0831.33.

[47] A. Abdullah, A. Siddique, K. Shaukat, and T. Jan, "An intelligent mechanism to detect multi-factor skin cancer," Diagnostics, vol. 14, no. 13, Art. no. 1359, 2024, doi: 10.3390/diagnostics14131359.

[48] K. Thurnhofer-Hemsi and E. Domínguez, "A convolutional neural network framework for accurate skin cancer detection," *Neural Process. Lett.*, vol. 53, no. 5, pp. 3073–3093, 2021, doi: 10.1007/s11063-020-10364-y.

[49] V. Ravi, "Attention Cost-Sensitive Deep Learning-Based Approach for Skin Cancer Detection and Classification," Cancers, vol. 14, no. 23, Art. no. 5872, Nov. 2022, doi: 10.3390/cancers14235872.

[50] A. A. Abdullah and M. Q. Hatem, "Audio transmission through Li-Fi," *International Journal of Civil Engineering and Technology*, vol. 9, no. 7, pp. 853–859, 2018, doi: 10.34218/IJCIET_09_07_088.

Author Biography

Wajid Dawood Alwan Al-Obaidi received his B.Sc. degree in Electrical Engineering from the University of Diyala, Iraq, in 2007, and his M.Sc. degree in Electronic Engineering from the University of Technology – Baghdad, Iraq, in 2014. Researcher Al-Obaidi is currently a Ph.D. student at the University of Babylon, College of Engineering, Department of Electronics and Communications Engineering. His research interests include medical image analysis, machine learning, deep learning, biomedical signal processing, and computer-aided diagnosis. He has published several papers in Scopus-indexed journals and actively contributes to research on hybrid deep-learning architectures and diagnostic models for healthcare applications. He can be contacted at: e-mail:eng162.wajed.dawood@student.uobabylon.edu.iq **ORCID: https://orcid.org/0009-0000-0364-1776**

Osama Qasim Jumah Al-Thahab was born in Al-Hilla City, Babil, Iraq, in 1978. He earned his B.Sc. in General Electrical Engineering from the University of Babylon in 2000, followed by his M.Sc. and Ph.D. in Electronics and Communications Engineering from the University of Technology, Iraq, in 2003 and 2007, respectively. Since 2019, he has been a Professor in the Department of Electrical Engineering at the University of Babylon. His research interests include microcontrollers, image processing, quadcopters, home automation, speech analysis, and digital design. He has published 19 articles in these fields. He can be contacted by email at eng.osama.qasim@uobabylon.edu.iq. **ORCID: 0000-0002-6033-0787.**

Hanaa Mohsin Ali Al Abboodi received her Ph.D. degree in Machine Learning and Data Science from the University of Salford, Manchester, UK, in 2019. She previously earned her M.Sc. in Information Technology from the University of Technology, Baghdad, Iraq, in 2005, and her B.Sc. in Computer Science in 1999. Her research interests include data science, machine learning, acoustic engineering, sound processing, HRTF, and image processing. Currently, she is a member of the teaching staff in the Department of Electrical Engineering at the University of Babylon. She has contributed to interdisciplinary research connecting computing, signal analysis, and engineering applications. Email: hanaa.ali@uobabylon.edu.iq; **ORCID: 0000-0003-3612-355.**

Manuscript received 14 May 2026; Revised 22 June 2026; Accepted 28 June 2026; Available online 6 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1757