Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Robust Non-Invasive Melanoma Early Detection Through Lightweight Self-Supervision and Lesion-Focused Representation Learning

Chen-Hao Peng
National Taiwan University
Tzu-Kun Lo
An-Nan Hospital
Da-Chuan Cheng
dccheng@mail.cmu.edu.tw

China Medical University

Article

Keywords: Melanoma early detection, lightweight deep learning, lesion-aware self-supervision, disease-specific masked autoencoder, robust mobile dermoscopy

Posted Date: September 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10789908/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Robust Non-Invasive Melanoma Early Detection Through Lightweight Self-Supervision and Lesion-Focused Representation Learning

Chen-Hao Peng[1,2], Tzu-Kun Lo[3,*], Da-Chuan Cheng[2,**]

[1] Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei, Taiwan.
[2] Department of Biomedical Imaging and Radiological Science, China Medical University, Taichung, Taiwan.
[3] Department of Dermatology, Tainan Municipal An-Nan Hospital, China Medical University, Tainan, Taiwan
[*] Corresponding Author: w_lee62@hotmail.com
[**] Corresponding Author: dccheng@mail.cmu.edu.tw

1. Abstract

Accurate and non-invasive classification of skin lesions is essential for early melanoma detection. However, many deep learning approaches depend on large external datasets, complex models and high-quality dermoscopic images. This reliance limits their use in mobile or resource-limited settings. We developed a lightweight YOLOv11-based framework that is trained only on the HAM10000 dataset and is designed to remain robust to degraded, smartphone-like images. The framework combines artifact removal, photometric and geometric augmentation for class balancing, and a hierarchical two-stage classifier that first distinguishes malignant from non-malignant lesions and then performs fine-grained categorization. We further integrate Shuffle Attention and a refined classification head to enhance lesion-focused feature extraction. In addition, we adapt a lesion-aware masked autoencoder strategy for self-supervised pretraining. Building upon existing region-aware masking concepts, it learns pathology-oriented representations by heavily masking lesion regions. Robustness is assessed with a stress test that applies bilateral filtering to simulate multiple levels of image degradation. Our method achieves 0.961 Top-1 accuracy and 0.999 AUC on HAM10000 without external data. Under strong simulated blur it maintains balanced accuracy above 0.930 and AUC of at least 0.997. These results indicate resilience to texture loss and common mobile imaging artifacts and support potential use in clinical workflows and teledermatology.

Keywords: Melanoma early detection, lightweight deep learning, lesion-aware self-supervision, disease-specific masked autoencoder, robust mobile dermoscopy

2. Introduction

Skin cancer is one of the most prevalent malignancies worldwide [1]. Its incidence continues to rise due to increasing ultraviolet exposure and an aging population, and it has become a major global public-health burden [1,2]. The World Health Organization estimates approximately 2–3 million new cases of non-melanoma skin cancer (NMSC) and about 132,000 new melanoma cases each year [3]. Although melanoma accounts for only 1–2% of skin cancers, it is highly aggressive. Without timely diagnosis and treatment, the five-year survival rate can fall below 30% [4]. NMSC occurs more frequently and usually progresses more slowly, but delayed detection can still cause severe tissue destruction and, in advanced cases, mortality [5]. Early detection of skin cancer is therefore critical for improving patient outcomes [6].

Clinical diagnosis mainly relies on dermatologist expertise, often assisted by dermoscopy, which provides detailed visualization of lesion structures [7,8]. Dermoscopy improves diagnostic accuracy, but outcomes still vary with physician experience [8,9]. Some lesion patterns remain difficult to interpret even for expert clinicians [10]. To address these issues, computer-aided diagnostic (CAD) tools have been developed to improve accuracy, consistency and diagnostic confidence [11]. In recent years, deep learning (DL) has achieved state-of-the-art performance in skin-lesion image analysis [7,12]. Many convolutional neural network (CNN)-based methods show strong results on public datasets [13–17], particularly through the integration of sophisticated feature extraction layers, multi-resolution pathways, and attention-guided mechanisms to capture both global and local lesion structures [18–21]. However, several limitations still hinder the use of these methods in routine clinical practice [22].

Most existing approaches adopt large-scale DL backbones such as ResNet, DenseNet and Vision Transformers (ViTs). These models can be highly accurate, but they contain many parameters and demand substantial computation, which limits deployment on mobile or edge devices [23]. Some methods perform lesion segmentation before classification to better focus on lesion regions. This two-step design increases system complexity and inference latency, which reduces practicality in clinical or mobile scenarios [24,25]. Lightweight architectures such as MobileNet and EfficientNet reduce model size and computation, but they may sacrifice accuracy, which is undesirable for high-stakes diagnostic tasks [26–28]. Thus, an important challenge is to achieve a good balance between accuracy and model efficiency, especially in resource-constrained environments.

In real-world use, patients may not notice subtle skin changes or seek medical attention for every small lesion. Automated systems therefore need to maintain high diagnostic accuracy on clinical dermoscopic images while also being robust to low-resolution, artifact-prone images

captured by smartphones [29,30]. Such robustness would broaden access to screening and support teledermatology and care in underserved regions [31,32]. Yet reliable adaptation to non-clinical, low-quality images remains insufficiently explored [33,34].

To address these challenges, we enhance YOLOv11 to build a lightweight yet high-performance framework for skin-lesion classification. We integrate an attention mechanism to refine lesion-centric feature learning and improve classification accuracy. We further optimize the classification architecture by adjusting the feature extraction layers and redesigning the classification head. These changes allow the model to better distinguish diverse skin lesions while reducing computational load and parameter count [35–37].

Current methods often require large labeled datasets and long training to reach satisfactory performance. To reduce this dependence, we integrate a region-aware self-supervised learning strategy based on Masked Autoencoders (MAE) [38,39]. By pretraining the model to reconstruct masked image regions, the network learns pathology-oriented and generalizable representations from limited data. This approach reduces reliance on extensive annotation and can accelerate convergence [40,41].

Inspired by existing malignancy-first classification pipelines, a key component of our framework is a two-stage hierarchical classification strategy. In the first stage, lesion images are grouped into four clinically aligned classes: one benign class and three malignant classes—melanoma, basal cell carcinoma and actinic keratoses and intraepithelial carcinoma. This stage prioritizes malignant-lesion recognition and mirrors real diagnostic workflows, where ruling out life-threatening disease is the main concern. In the second stage, we apply transfer learning to further classify the seven lesion categories defined in the HAM10000 dataset. This provides fine-grained differentiation once malignancy has been assessed. The staged design improves the reliability of the initial benign–malignant decision and enhances subtype discrimination, while adding minimal computational and training overhead.

To evaluate real-world practicality, we also perform robustness testing using bilateral filtering to simulate low-quality smartphone imaging and different levels of texture degradation. This setup allows us to systematically examine performance under realistic non-clinical conditions, where blur and artifacts are common. Overall, our goal is to improve skin-lesion classification accuracy while preserving computational efficiency and avoiding large increases in model size. By enhancing YOLOv11 with lesion-aware attention, self-supervised representation learning and a clinically motivated hierarchical classifier, we aim to provide a lightweight model suitable for mobile deployment. This capability could facilitate broader adoption of early skin-cancer detection in clinical workflows and teledermatology.

The main contributions of this study are as follows:

1. **Lightweight high-performance framework:** We develop a YOLOv11-based skin-lesion classification framework that is optimized for resource-constrained deployment while preserving state-of-the-art accuracy and robustness to low-quality mobile imaging.
2. **Lesion-focused architecture:** We integrate Shuffle Attention and a structured classification head to strengthen lesion-centric feature representation, improving discrimination with minimal computational overhead.
3. **Lesion-aware self-supervision:** We adapt a region-aware self-supervised pretraining strategy with a Masked Autoencoder, reducing reliance on large labeled datasets and accelerating training.
4. **Clinically motivated hierarchical strategy:** We implement an integrated two-stage hierarchical classification workflow based on malignancy-first screening principles that prioritizes malignant-lesion screening before fine-grained categorization via transfer learning.
5. **System-level integration and empirical evaluation**: We provide a comprehensive empirical evaluation of integrating artifact removal, region-aware self-supervision, attention mechanisms, and hierarchical training into a single optimized pipeline for mobile deployment.

3. Materials and Methods

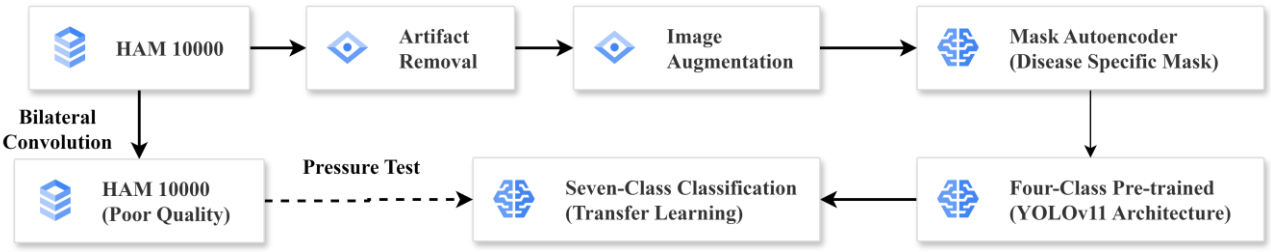


Figure 1 illustrates the overall pipeline. Dermoscopic images from HAM10000 are first preprocessed to remove artifacts and then augmented to increase class balance and diversity. The enhanced dataset is used to train a four-class malignant-prioritized classifier, after which transfer learning is applied for seven-class fine-grained classification.

3.1 Dataset

The Human Against Machine with 10,000 Training Images (HAM10000) dataset is a dermoscopic benchmark for skin-lesion classification published by Tschandl et al. in 2018 [9]. It contains 10,015 dermoscopic images collected from multiple clinical sources, including dermatology clinics in Austria and Germany. All images were acquired under standardized

dermoscopy protocols and stored as JPEG files with a resolution of 600 × 450 pixels. Each image is accompanied by expert annotations specifying lesion type, image source and, for some cases, anatomical site. Diagnoses were confirmed using histopathology, dermoscopic consensus and/or long-term follow-up, providing reliable ground truth. An overview of the full data-processing and training pipeline based on the HAM10000 images is shown in Figure 1.

HAM10000 includes seven lesion categories (Table 1). The nevus (NV) class is the largest and accounts for roughly 67% of all images. Dermatofibroma (DF) and vascular lesions (VASC) are the smallest classes, which creates substantial class imbalance. The dataset also spans diverse patient demographics and acquisition settings, resulting in natural variation in illumination, color and skin tone. Because of these characteristics, HAM10000 is widely used to train and evaluate automated skin-lesion classification models.

Table 1. Categories and image distribution in the HAM10000 dataset

Disease Name	Abbreviation	Number of Images	Lesion Type
Melanoma	MEL	1113	Malignant
Basal Cell Carcinoma	BCC	514	Malignant
Actinic Keratoses and Intraepithelial Carcinoma (including Squamous Cell Carcinoma and Bowen's Disease)	AKIEC	327	Malignant
Nevus	NV	6705	Benign
Benign Keratosis (Including Seborrheic Keratosis)	BKL	1099	Benign
Dermatofibroma	DF	115	Benign
Vascular Lesions (e.g., Hemangioma)	VASC	142	Benign

To evaluate the model and prevent data leakage, we implemented a data splitting protocol. First, to ensure patient-level separation, we utilized the lesion_id metadata provided in the dataset. All images belonging to the same physical lesion were grouped and assigned exclusively to either the training or the validation set, preventing the model from recognizing redundant spatial features of the same lesion across different sets. The partition followed the official ISIC 2018 Challenge Task 3 protocol [13], under which the 10,015 HAM10000 images constitute the training corpus and a further 193 validation and 1,512 held-out test images are provided separately. All results reported below were therefore obtained on the official held-out test set rather than on a subset carved out of the training data. Furthermore, a strict boundary was maintained between data augmentation and robustness testing. All data augmentation procedures were restricted entirely to the training set. The validation and test sets comprised only original, unmodified images to reflect true clinical inference conditions. Conversely, the bilateral filtering procedure, which was

designed as a pressure test to simulate low-quality smartphone imaging, was applied exclusively to the test set during the final evaluation phase. The model was never exposed to these degraded images during training or validation. Finally, all hyperparameter tuning and checkpoint selections were conducted based strictly on the validation set.

To evaluate the generalizability and robustness of the proposed framework on independent clinical data, we utilized the International Skin Imaging Collaboration (ISIC) 2019 dataset as an external validation set. Since the HAM10000 dataset is a known subset of ISIC 2019, we implemented the deduplication protocol established by Cassidy et al. [42] to identify and remove all overlapping images and near-duplicates between our training set and the external test set. This filtering ensured that the model was evaluated on truly unseen samples, thereby eliminating potential performance inflation due to data leakage. The resulting independent test set comprised diverse lesion categories with class distributions similar to the primary training set, providing a benchmark for assessing the model's adaptability to varied acquisition environments and populations.

3.2 Artifact Removal

Dermoscopy images often contain non-lesion artifacts—such as hair, shadows, and specular reflections—that obscure lesion boundaries and can mislead feature extraction. To mitigate this, we designed a multi-step preprocessing pipeline focused on hair removal and lesion restoration.

First, a black-hat morphological transformation was applied to enhance dark, thin hair structures against brighter skin backgrounds [43]. To improve sensitivity to curved and elongated patterns, we constructed a 13 × 13 diagonal cross-shaped kernel with small gaps (Figure S1), which accentuates hair-like features while preserving lesion morphology.

$$I_{bh} = (I \bullet K_{13\times13}) - I \tag{1}$$

where I is the raw image, $K_{13\times13}$ is the diagonal cross-shaped kernel (Figure S1), $\bullet$ denotes morphological closing, and I_{bh} is the resultant image.

Next, Otsu's thresholding generated a threshold t^* (Eq. (2)) by separating artifact (hairs/shadows/reflections) pixels from background based on minimum intra-class variance on image I_{bh} and result an output image I_b. Another structural element using a 9 × 9 cross-shaped was applied on I_b with morphological closing operation to connect fragmented artifact segments and remove small discontinuities. This is used to detect artifact regions: $I_{artifact} = (I_b \oplus K_{9\times9}) \ominus K_{9\times9}$. Finally, bicubic interpolation [44] was used to inpaint the artifact regions ($I_{artifact}$), reconstructing occluded lesion areas with smooth, artifact-free surfaces while maintaining natural skin texture.

$$t^* = \arg\max_t \sigma_B^2(t) \tag{2a}$$

$$\sigma_B^2 = p_1(\mu_1 - \mu)^2 + p_2(\mu_2 - \mu)^2 = \frac{p_1}{1 - p_1}(\mu - \mu_1)^2 \quad (2b)$$

where p_1, p_2 are class probabilities on histogram of image I_{bh}, and μ, μ_1 are image mean and class 1 mean under threshold t^*.

This preprocessing produces cleaner and more consistent dermoscopic images, allowing the model to focus on diagnostically meaningful lesion cues rather than nuisance artifacts.

3.3 Data Augmentation

To mitigate the pronounced class imbalance in the HAM10000 dataset, we adopted a targeted augmentation scheme to expand minority lesion categories. Images from underrepresented classes were augmented until each category reached the same size as the dominant nevus (NV) class. Formally, the target number of samples per class was defined as: $N_{target} = N_{NV} = 6705$, and for each class *c*, the required augmentation multiplier was

$$k_c = \left\lceil \frac{N_{target}}{N_c} \right\rceil, \; N_c' = k_c N_c \approx N_{target} \quad (3)$$

where N_c is the original number of images in class *c*, k_c is the class-specific augmentation factor, and N_c' is the resulting balanced class size. This balancing step reduces bias toward majority classes, stabilizes learning, and improves generalization across diverse lesion morphologies.

Augmentation was implemented using the Albumentations library [45] and consisted of both geometric and photometric transformations. Geometric operations were designed to reflect realistic variations in lesion orientation and shape, including horizontal and vertical flips, random rotations within ±30°, affine transformations, perspective warping, and elastic deformation. These transformations emulate changes in viewing angle, lesion positioning, and mild skin stretching that commonly occur during dermoscopic acquisition. Photometric transformations were applied to represent real-world variability in illumination and device-dependent color response, including hue–saturation shifts, brightness and contrast adjustments, gamma correction, and mild motion blur.

For each training image, two to four transformations were randomly sampled and sequentially composed to generate an augmented instance. For a case with four transformations, we obtain the augmentation as follows:

$$\tilde{I} = \mathcal{T}_4 \circ \mathcal{T}_3 \circ \mathcal{T}_2 \circ \mathcal{T}_1(I), \quad (4)$$

where *I* was the original image, $\tilde{I}$ was the augmented image, and $\mathcal{T}_i$ denoted a randomly selected geometric or photometric transform. This randomized composition yields structurally and chromatically diverse yet diagnostically consistent samples.

Representative augmented examples were shown in Figure S2. The procedure preserved clinically relevant cues—such as lesion boundaries, pigment networks, and global color distribution—while introducing controlled variability in geometry and illumination. Overall, this

augmentation strategy produced a balanced and richer training set, strengthens robustness to acquisition heterogeneity, and encouraged the model to learn lesion-discriminative features rather than class-frequency priors.

3.4 Model Architecture and Optimization Strategy

To build a fast and lightweight diagnostic framework, we adopted YOLOv11 [46] as the backbone, leveraging its favorable speed–accuracy trade-off for real-time medical imaging. As shown in Figure 2, the overall architecture comprises convolutional feature extractors, lightweight attention modules, and a structured classification head tailored for skin-lesion discrimination.

Disease-focus MAE pretraining. To enhance representation learning under limited data, we introduced a disease-specific masked autoencoder (DS-MAE) for self-supervised pretraining within the YOLOv11 backbone. Unlike standard MAE approaches that randomly masked image patches, DS-MAE used automatically generated lesion masks to guide occlusion. To ensure reproducibility without manual annotation, these masks were deterministically generated using an automated morphological pipeline. Input images were converted to grayscale, enhanced via Contrast Limited Adaptive Histogram Equalization (CLAHE, clip limit=2.0, 8×8 grid), inverted, and binarized using Otsu's thresholding. The masks were further refined with a median blur (kernel size=5) and morphological closing and opening (using a 5×5 elliptical structuring element), followed by a connected component analysis to isolate the primary lesion region. Let Ω_L denote the lesion region and $u(p)\sim\mathrm{U}(0,1)$ be a uniform random variable for pixel p. The lesion-aware masking rule was defined as

$$\mathrm{M(p)} = \begin{cases} 1, & p \in \Omega_L \text{ and } u(p) < 0.8 \\ 1, & p \notin \Omega_L \text{ and } u(p) < 0.1 \\ 0, & \text{otherwise} \end{cases}, \tag{5}$$

where $\mathrm{M(p)}=1$ indicated that pixel p is masked. Specifically, 80% of the pixels within the lesion mask and 10% of the background pixels were selected and replaced with zero values (black pixels). Empirical sensitivity analyses confirmed this ratio to be optimal; masking over 90% of the lesion severely degraded the decoder's reconstruction stability, whereas masking less than 50% failed to effectively force the network to extract deep, pathology-relevant structural features. This strategy forced the encoder to reconstruct clinically relevant lesion morphology, chromatic patterns, and boundaries, thereby promoting pathology-oriented and generalizable representations. To maintain computational efficiency aligned with our lightweight objective, the DS-MAE employed a fully convolutional decoder rather than a heavier Transformer-based decoder. This decoder consisted of a series of transposed convolutional layers (ConvTranspose2d) paired with Batch Normalization and ReLU activations to progressively upsample the latent features, terminating

with a Sigmoid activation. Reconstruction was optimized using the mean squared error (MSE) loss over masked pixels:

$$\mathcal{L}_{MAE} = \frac{1}{|\Omega_M|}\sum_{p\in\Omega_M}\left\|I(p) - \hat{I}(p)\right\|_2^2,\ \Omega_M = \{p \mid M(p) = 1\} \quad (6)$$

where I and $\hat{I}$ denoted the original and reconstructed images, respectively, and Ω_M was the masked pixel set. Adadelta [47] was used during pretraining together with a cyclical learning-rate schedule [48] ranging from 10^{-5} to 10^{-1} to stabilize convergence and reduced the risk of local minima. The self-supervised pretraining was conducted using a batch size of 16.

Lesion-focused attention and classification head. For efficient lesion emphasis, Shuffle Attention [49] was inserted before the fully connected layers. Given an intermediate feature map $F \in \mathbb{R}^{C\times H\times W}$, Shuffle Attention produced a reweighted feature map F' by combining channel and spatial attention with lightweight shuffling:

$$F' = \mathrm{SA}(F) = \alpha(F)\odot F, \quad (7)$$

where $\alpha(F)$ was the learned attention map and $\odot$ denoted element-wise multiplication. This module enhanced lesion-centric cues with minimal additional computation. We further redesigned the classification head using stacked dropout, batch normalization, and ReLU layers to progressively refine high-dimensional features, mitigate overfitting, and preserved lesion-specific information.

Two-stage hierarchical training. Training proceeded in two stages. In the first stage, the model performed coarse four-class classification (melanoma, basal cell carcinoma, actinic keratoses and intraepithelial carcinoma, and benign) to prioritize malignant detection. In the second stage, transfer learning was applied for seven-class HAM10000 classification by retaining the pretrained YOLOv11 backbone and replacing only the final classification layer, enabling efficient adaptation without retraining the full network. For each stage, classification was optimized using the cross-entropy loss:

$$\mathcal{L}_{cls} = -\sum_{c=1}^{C} y_c \log \hat{y}_c, \quad (8)$$

where $\mathrm{C} = 4$ in stage one and $\mathrm{C} = 7$ in stage two, y_c was the one-hot ground truth label, and $\hat{y}_c$ was the predicted probability for class c. Both stages were optimized with the Lion optimizer [50]. ($\text{learning rate } 0.001, \beta_1 = 0.9, \beta_2 = 0.99$) In brief, Lion interpolated between the momentum and the current gradient and applied a single sign operation to the result:

$$c_t = \beta_1 m_{t-1} + (1-\beta_1) g_t, \theta_t = \theta_{t-1} - \eta(\text{sign}(c_t) + \lambda\theta_{t-1}), m_t = \beta_2 m_{t-1} + (1-\beta_2) g_t \quad (9)$$

where $g_t = \nabla_\theta \mathcal{L}$ was the gradient, m_t was the momentum estimate, η was the learning rate, c_t was the interpolated update direction to which the sign operation was applied, and λ was the decoupled weight-decay coefficient. Weight decay was applied selectively: primary convolution weights received no decay, regularized layers used 5×10^{-4}, and bias parameters were excluded. The complete training workflow was summarized in Figure 2.

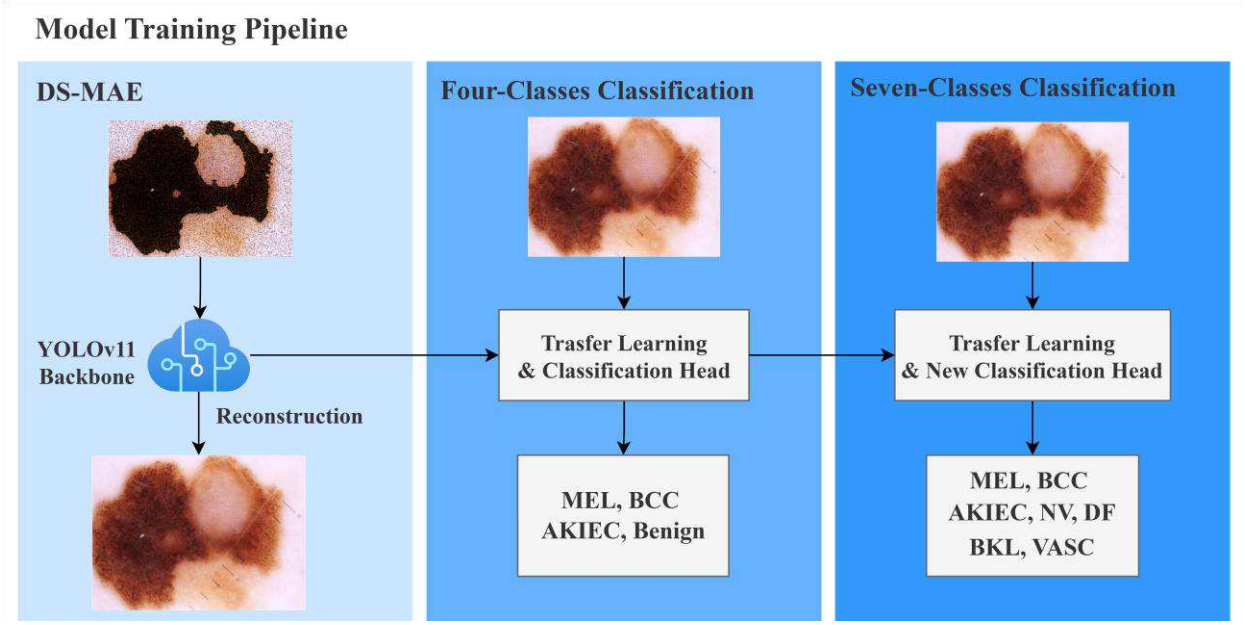


Figure 2. Overview of the proposed workflow. The YOLOv11 backbone was first pretrained using a DS-MAE with lesion-aware reconstruction to learn pathology-oriented representations. The pretrained network was then trained in a two-stage hierarchy: a four-class malignant-prioritized coarse classifier followed by transfer-learning based fine-tuning for seven-class HAM10000 lesion classification.

3.5 Low-Quality Image Simulation Using Bilateral Filtering

To assess the robustness of the proposed framework under realistic mobile conditions, we simulated low-quality dermoscopic images resembling those captured by smartphones or consumer-grade cameras. We adopted bilateral filtering [51] to intentionally reduce image sharpness while preserving diagnostically important lesion boundaries. Unlike Gaussian blurring, which smoothed all regions uniformly, bilateral filtering performed edge-preserving smoothing by jointly considering spatial proximity and pixel-intensity similarity. This property enabled controlled texture attenuation that mimicked optical defocus or motion-induced degradation without eliminating critical structural cues.

Formally, given an input image I, the bilateral-filtered output I' at pixel location p is

computed as:

$$\mathrm{I}'(\mathrm{p}) = \frac{1}{w_p}\sum_{q\in N(p)} G_{\sigma_s}(\|p-q\|)\, G_{\sigma_r}(|I(p)-I(q)|)I(q), \tag{10}$$

where $N(p)$ denoted a local neighborhood around p, σ_s controled the degree of spatial smoothing, and σ_r controled intensity-domain smoothing. $G_{\sigma_s}(\|p-q\|)$ was the spatial Gaussian kernel, which assigned higher weights to nearby pixels. $G_{\sigma_r}(|I(p)-I(q)|)$ was the range Gaussian kernel, which suppressed contributions from pixels with dissimilar intensities, thereby preserving edges. The w_p was the normalization factor to ensure the weights sum to one, calculated as the sum of all the weights in the numerator in (10).

To create progressive degradation levels, we gradually increased the kernel size and the smoothing parameters (σ_s, σ_r), generating two representative blur settings: mild and strong. As the filter strength increased, fine lesion textures and subtle color gradients were progressively smoothed, closely reproducing the visual artifacts commonly observed in non-clinical acquisition environments (e.g., defocus blur, motion blur, and phone-camera noise). This controlled degradation protocol enables systematic evaluation of the model's stability and adaptability under suboptimal imaging conditions. Consequently, performance could be assessed not only on high-quality dermoscopic images but also on realistically blurred inputs, supporting the framework's applicability to teledermatology and real-world mobile screening.

3.6 Model Evaluation

Given the substantial class imbalance in HAM10000, we evaluated performance using a set of complementary metrics: Accuracy (ACC), Balanced Accuracy (BA), Precision, F1-score, Matthews Correlation Coefficient (MCC), Jaccard Score (JS), and one-vs-one Area Under the ROC Curve (OvO AUC). ACC provided an overall correctness measure, whereas BA reduced bias introduced by the dominant nevus (NV) class by averaging recall across categories. Per-class as well as weighted Precision and F1-score were reported to characterize sensitivity and specificity for individual lesion types. MCC and JS further quantified classification quality by accounting for both correct and incorrect predictions, which was particularly informative for minority classes such as DF and VASC. Finally, we adopted OvO AUC instead of the one-vs-all (OvA) variant because it evaluated separability between every pair of lesion types, yielding a stricter assessment of multi-class discrimination. Together, these metrics provided a comprehensive and unbiased evaluation of diagnostic performance.

Let C be the number of classes and let $TP_c, FP_c, FN_c,$ and TN_c denote the true positives, false positives, false negatives, and true negatives for class c, respectively. The overall accuracy was

$$\mathrm{ACC} = \frac{\sum_{c=1}^{C} TP_c}{N} \tag{11a}$$

where N was the total number of test images. Balanced accuracy was computed as the mean class

recall:

$$\mathrm{BA} = \frac{1}{C}\sum_{c=1}^{C}\frac{TP_c}{TP_c+FN_c} \quad . \tag{11b}$$

For each class, Precision, Recall, and F1-score are defined as:

$$\mathrm{Precision}_c = \frac{TP_c}{TP_c+FP_c} \tag{11c}$$

$$\mathrm{Recall}_c = \frac{TP_c}{TP_c+FN_c} \tag{11d}$$

$$\mathrm{F1}_c = \frac{2\mathrm{Precision}_c\mathrm{Recall}_c}{\mathrm{Precision}_c+\mathrm{Recall}_c} \tag{11e}$$

We further reported the Jaccard Score (intersection-over-union) per class:

$$\mathrm{JS}_c = \frac{TP_c}{TP_c+FP_c+FN_c} \tag{11f}$$

To provide a correlation-based measure robust to imbalance, we computed MCC using the confusion matrix $K \in \mathbb{R}^{C\times C}$, where is the number of samples of true class i predicted as class j. Let and denote the number of samples whose true class is i and the number of samples predicted as class i, respectively. The multi-class MCC is:K_{ij}

$$\mathrm{MCC} = \frac{\sum_i \sum_j \sum_k K_{ii}K_{jk}-K_{ij}K_{ki}}{\sqrt{(\sum_i T_i)^2-\sum_i T_i^2}\ \sqrt{(\sum_i P_i)^2-\sum_i P_i^2}} \tag{12}$$

4. Results

To evaluate the effectiveness of the preprocessing pipeline, we first examined artifact removal on dermoscopic images. The pipeline consistently suppressed hair artifacts and other thin linear structures while preserving native skin texture and colour. As shown in Figure S3, the black-hat morphological operation enhanced dark hair strands against the brighter background and enabled accurate foreground segmentation using Otsu thresholding. The resulting binary mask clearly separated hair regions from lesion tissue and surrounding skin, providing a precise estimate of occluded areas.

Subsequent inpainting generated reconstructed images with smooth surfaces and a markedly reduced artifact burden, while preserving lesion morphology and boundary integrity. This restoration improved both the visual quality and structural coherence of the dermoscopic images and produced cleaner inputs for representation learning. In practice, the enhanced continuity and reduced clutter allowed the model to focus on diagnostically relevant patterns rather than spurious non-lesion structures, thereby supporting more reliable feature extraction and classification.

To illustrate the pretraining behaviour of the proposed DS-MAE, we examined examples of lesion-aware masking on dermoscopic images (Figure S4). Unlike conventional random masking, DS-MAE uses lesion segmentation maps to guide patch occlusion and directs the model's attention to clinically relevant regions. During pretraining, 80% of pixels within the lesion area and 10% of pixels in the surrounding background are masked.

This selective masking forces the encoder to learn discriminative, pathology-oriented representations by focusing on lesion morphology, colour patterns and boundary structures. As shown in Figure S4, the lesion-focused masking preserves global spatial context but deliberately challenges the model to recover diagnostically important features. This process strengthens the quality of the learned features and benefits downstream classification performance.

We performed an ablation study to quantify the contribution of each training component to the final classification performance. Table S1 summarizes the results and shows a clear, progressive improvement as we added preprocessing, data augmentation, DS-MAE pretraining, Shuffle Attention and the structured classification head. To ensure fair benchmarking and reproducibility, all baseline and comparative models in this ablation study were evaluated under identical hyperparameter settings and hardware environments.

The baseline YOLOv11s model achieved a Top-1 accuracy of 0.880 and a balanced accuracy of 0.785. Adding the preprocessing pipeline slightly increased performance to 0.881 Top-1 accuracy and 0.808 balanced accuracy, reflecting the benefit of artifact reduction. Incorporating data augmentation further improved generalization and increased balanced accuracy to 0.826.

A substantial gain was observed after DS-MAE pretraining, which raised balanced accuracy to 0.923 and markedly strengthened feature representations. Adding four-class malignant-prioritized pretraining led to additional improvements, yielding 0.939 Top-1 accuracy and an MCC of 0.902. The final optimized configuration, which combined all modules, achieved the best performance: 0.961 Top-1 accuracy, 0.971 balanced accuracy, an MCC of 0.935 and a OvO AUC of 0.999, indicating near-perfect multi-class discrimination.

Table 2 summarizes the classification performance of the best-performing model. The model achieved a Top-1 accuracy of 0.961, indicating reliable overall classification across all lesion types. The high balanced accuracy further confirmed consistent performance on both majority and minority classes. An OvO AUC of 0.999 reflected excellent discrimination across all categories.

At the per-class level, the model achieved near-perfect classification for AKIEC and VASC, with MCC and F1-scores of 0.999, indicating that virtually no samples were misclassified. BCC also showed near-perfect performance, with 0.999 accuracy and an OvO AUC of 0.999. For BKL,

the model reached a Top-1 accuracy of 0.981, but the balanced accuracy was slightly lower at 0.943, suggesting some confusion with visually similar lesion types. MEL showed a modest reduction in MCC and F1-score, indicating that a small number of melanoma samples were misclassified. We further examined this behaviour using the confusion matrix to identify inter-class overlaps.

Table 2. Overall classification performance of the best-performing model.

Diseases	Top-1 Accuracy	Balanced Accuracy	MCC	Jaccard Score	F1-score	Precision	OvO AUC
AKIEC	0.999	0.999	0.999	0.999	0.999	0.999	0.999
BCC	0.999	0.999	0.994	0.989	0.995	0.989	0.999
BKL	0.981	0.943	0.923	0.869	0.931	0.975	0.999
DF	0.999	0.989	0.988	0.977	0.989	0.999	0.999
MEL	0.976	0.963	0.885	0.814	0.898	0.853	0.999
NV	0.967	0.966	0.931	0.946	0.972	0.972	0.999
VASC	0.999	0.999	0.999	0.999	0.999	0.999	0.999
Overall	0.961	0.971	0.935	0.942	0.961	0.963	0.999

To further evaluate classification performance, we analysed both absolute and normalized confusion matrices (Figure 3). In the absolute matrix, the model correctly classified all samples of AKIEC, BCC and VASC, indicating highly reliable recognition of these categories. The NV class had the highest number of correctly identified samples, with 884 true positives. However, a small number of MEL and BKL cases were incorrectly predicted as NV. For MEL, eight cases were classified as NV and one as BKL, revealing partial overlap between these visually similar lesions.

The normalized confusion matrix highlighted these trends more clearly, where the values along the main diagonal represent the recall for each class. AKIEC, BCC and VASC all achieved a recall of 0.999 with virtually no false negatives, confirming stable classification for these classes. BKL reached a recall of 0.893, with about 7% of cases misclassified as NV. MEL attained a recall of 0.951, with around 5% of cases predicted as NV. These findings show that, although the model performs exceptionally well for most lesion types, some confusion remains between BKL, MEL and NV, likely due to similar pigmentation and textural patterns.

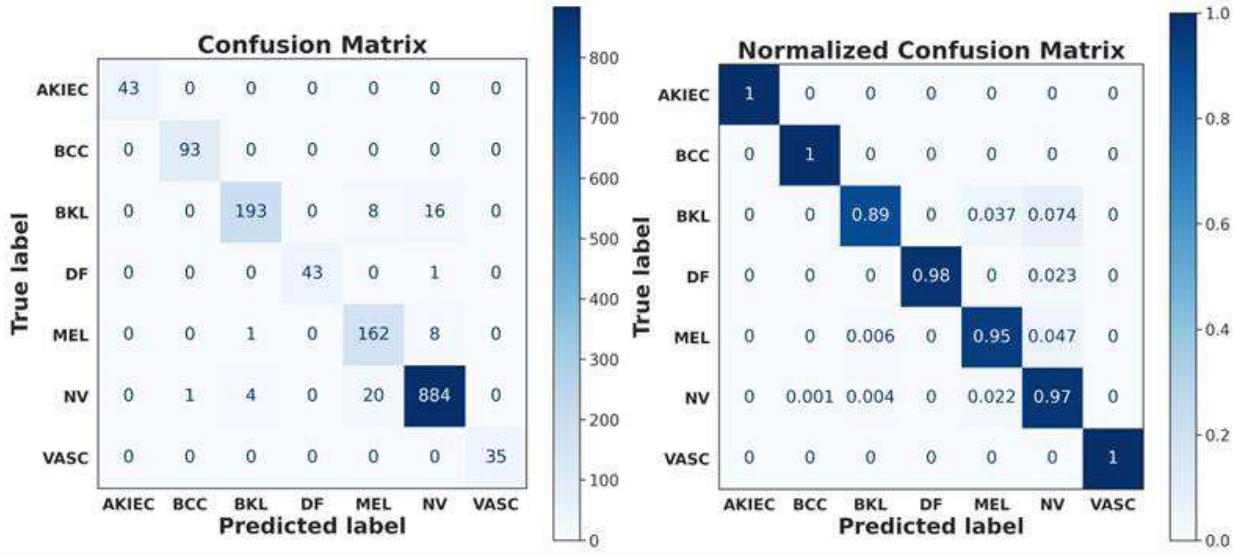


Figure 3. Confusion matrices of the best-performing model. (a) Absolute confusion matrix (b) Normalized confusion matrix

To simulate real-world imaging variability, we generated progressively blurred dermoscopic images using bilateral filtering at different intensity levels. As shown in Figure S5, this process mimicked the visual degradation commonly observed in low-quality or out-of-focus dermoscopic and smartphone-captured images. The resulting images exhibited a gradual loss of fine lesion texture and contrast while preserving overall structure and color distribution. This pressure test provided a visual benchmark for assessing the model's robustness under varying image-quality conditions.

Table 3. Classification performance under different image quality conditions.

Image Quality	Top-1 Accuracy	Balanced Accuracy	MCC	Jaccard Score	F1-score	Precision	OvO AUC
Original	0.961	0.971	0.935	0.942	0.961	0.963	0.999
Mild Bilateral Smoothing	0.951	0.967	0.921	0.922	0.952	0.956	0.999
Strong Bilateral Smoothing	0.849	0.934	0.787	0.768	0.860	0.900	0.997

Table 3 summarizes the model's performance under different image-quality conditions generated by bilateral filtering. The model showed strong robustness to mild blurring. Top-1 accuracy decreased only slightly from 0.961 to 0.951, and balanced accuracy remained above 0.960 under mild blurring. These results indicate that the classifier continued to provide reliable lesion discrimination despite moderate texture loss. Under strong blurring, where fine lesion

structures were heavily degraded, performance declined more noticeably, particularly for MCC and Jaccard score, reflecting increased confusion among visually similar classes. However, the OvO AUC remained at or above 0.997 across all conditions. This finding suggests that the model preserved a high level of discriminative capability even on severely degraded images. Overall, the framework appears resilient to realistic variations in the quality of dermoscopic and smartphone-acquired images.

Table 4. Classification performance on the independent ISIC 2019 external validation dataset after deduplication.

Diseases	Top-1 Accuracy	Balanced Accuracy	MCC	Jaccard Score	F1-score	Precision	OvO AUC
AKIEC	0.993	0.963	0.886	0.800	0.889	0.851	0.999
BCC	0.993	0.966	0.942	0.897	0.946	0.956	0.999
BKL	0.965	0.924	0.855	0.778	0.875	0.894	0.992
DF	0.997	0.966	0.952	0.911	0.953	0.976	0.999
MEL	0.946	0.916	0.762	0.649	0.787	0.714	0.992
NV	0.941	0.942	0.878	0.904	0.951	0.966	0.994
VASC	0.999	0.999	0.999	0.999	0.999	0.999	0.999
Overall	0.918	0.924	0.866	0.849	0.924	0.924	0.996

Table 4 presents the classification performance of the proposed model on the ISIC 2019 external validation dataset. To prevent data leakage, overlapping images with the HAM10000 training set were removed prior to evaluation. Under this independent testing condition, the model achieved an overall Top-1 accuracy of 0.918, a balanced accuracy of 0.924, and an OvO AUC of 0.996. At the class level, the model demonstrated robust performance across all seven lesion categories, maintaining precision and F1-scores above 0.850 for most classes. These quantitative metrics confirm that the classification capabilities learned from the primary dataset are maintained when applied to an external, unseen dataset.

To place our results in context, we compared the proposed model with state-of-the-art skin-lesion classification methods from previous studies. Table 5 reports the performance of our best model alongside FixCaps [52], IRv2 + Soft Attention [53], Multi-Resolution EfficientNets [54], a fuzzy rank-based deep ensemble [55], EFAM-Net [56], DeepSkin [57] and a YOLOv8-based classifier [58]. YOLOv11s with all proposed enhancements achieved a Top-1 accuracy of 0.961, which is close to FixCaps, the best-performing reference method, at 0.965. Our model also reached an OvO AUC of 0.999, exceeding the 0.984 reported for IRv2 + Soft Attention and demonstrating superior discrimination capability. In contrast to several of the reference methods, which drew on training data beyond HAM10000, our method maintained high accuracy using

only the HAM10000 dataset. The evaluation protocols behind these figures also differ. Several of the reference methods report accuracy on a subset held out from the challenge training corpus, whereas the present results were obtained on the official ISIC 2018 held-out test set, which is not available to the model at any point during development. Accuracies obtained under the former protocol are therefore not strictly comparable with ours.

Table 5. Performance comparison between our model and prior approaches.

Model	Accuracy	AUC	Average Precision	Extra Training Data	Year
YOLOv11s + Full Enhancements (ours)	0.961	0.999	0.963	No	2025
FixCaps [52]	0.965	-	-	Yes	2022
Fuzzy Rank-Based Deep Ensemble [55]	0.951	-	-	No	2025
EFAM-Net [56]	0.940	-	-	No	2024
IRv2+Soft Attention [53]	0.934	0.984	0.937	Yes	2021
Multi-Resolution Efficient Nets [54]	0.926	-	-	Yes	2020
DeepSkin [57]	0.912	-	-	No	2023
YOLOv8x-cls [58]	0.862	-	-	No	2024

The results were further visualized in Figure 4, where we compared the accuracy of different models and indicated whether additional training data was used. As shown in Figure 4, our model ranked among the top-performing methods while maintaining strong generalization capability without relying on external datasets.

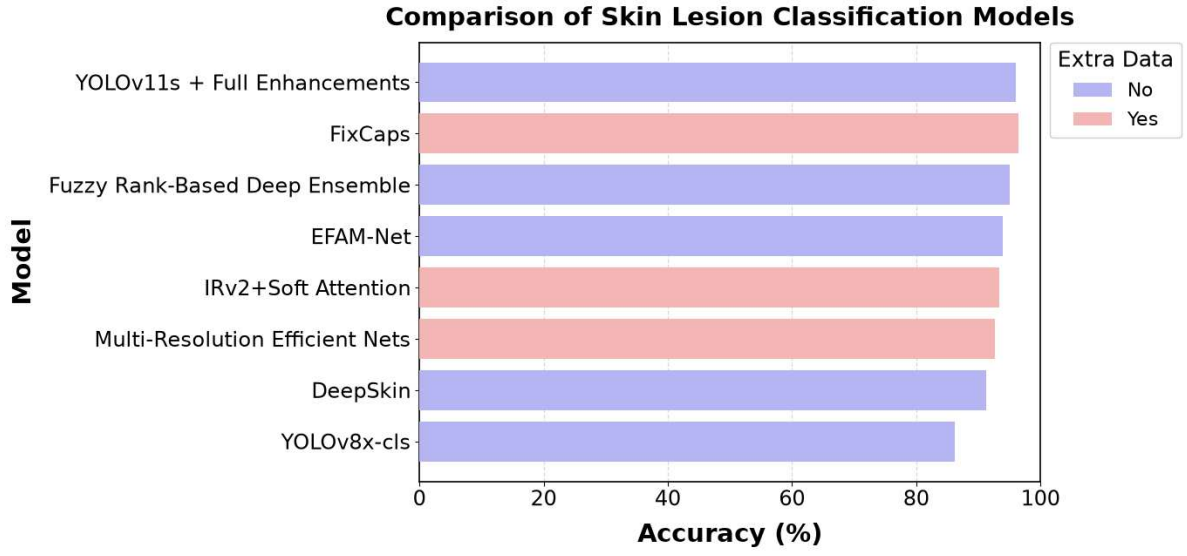


Figure 4. Accuracy comparison of different skin lesion classification models.

5. Discussion

This study aimed to establish a robust and lightweight diagnostic framework by integrating artifact removal, lesion-aware self-supervision through DS-MAE, and a two-stage hierarchical training strategy. Our results demonstrate that state-of-the-art performance can be achieved using exclusively the HAM10000 dataset. This mitigates the common reliance on massive external training data, which is often inaccessible in clinical practice. The proposed model not only reached a Top-1 accuracy of 0.961 but also achieved a superior OvO AUC of 0.999, indicating exceptionally strong discriminative power across all lesion categories. These figures were obtained on the official ISIC 2018 held-out test set rather than on a partition carved out of the training corpus, so they reflect performance on data that remained entirely unseen throughout model development.

The progressive performance gains observed in our ablation study highlight the synergistic effect of each methodological component. Preprocessing effectively neutralized confounding structural noise, such as hair and illumination artifacts, which frequently mislead feature extraction in dermoscopic analysis. The two-stage hierarchical training strategy further addressed the critical challenge of class imbalance. By first prioritizing the distinction between malignant and benign categories before attempting fine-grained subtyping, the model learned to disentangle broad pathology-relevant features from subtle inter-class variations. This structured approach mirrors clinical triage workflows and significantly contributed to the substantial increase in balanced accuracy compared to standard single-stage training.

A key component of our proposed framework is the adaptation of DS-MAE for self-supervised pretraining. Unlike conventional random masking, our lesion-aware masking strategy occludes 80% of the lesion while preserving most of the background. This approach forces the network to reconstruct complex diagnostic cues such as chromatic patterns and boundary morphology. The inclusion of DS-MAE clearly enhanced the robustness of feature representations, allowing the model to maintain higher MCC and Jaccard scores even in the presence of high intra-class variability. These findings suggest that self-supervised and lesion-centric reconstruction can effectively reduce dependence on large annotated datasets, which is particularly valuable in dermatology where expert labeling is resource-intensive.

Beyond diagnostic accuracy, the computational efficiency of our model provides a distinct advantage for real-world deployment. Compared to heavier architectures like EfficientNet V2 B1 or MobileViT V2, our YOLOv11-based framework utilizes significantly fewer parameters at comparable GFLOPs, while achieving the fastest GPU inference time of 4.88 ms and maintaining viable CPU latency. This favorable balance between model size and speed underscores the

suitability of the framework for edge computing and mobile teledermatology, where real-time feedback is essential. Furthermore, the pressure tests using bilateral filtering confirmed the resilience of the model. The retention of high diagnostic separability under severe image degradation suggests that our DS-MAE-driven features focus on global lesion morphology rather than pixel-level sharpness, making the system robust against defocus artifacts typical of smartphone photography.

While the overall classification performance is robust, a critical evaluation of misclassifications is essential for clinical safety. In melanoma screening, false negatives represent the most severe type of diagnostic error. This occurs when a malignant melanoma is misdiagnosed as a benign nevus or keratosis. As indicated by our confusion matrix, 9 melanoma cases were misclassified as benign lesions, including 8 as NV and 1 as BKL. Morphologically, these errors typically occur in early-stage or featureless melanomas that lack classic malignant indicators like atypical networks or blue-white veils, making them visually mimic atypical nevi. Clinically, such false negatives could lead to delayed treatment and compromised patient survival. Therefore, our framework is designed to function as an assistive triage tool rather than an autonomous diagnostician, serving as a second reader to alert clinicians to subtle malignant features.

Despite these advancements, certain limitations must be addressed. While the model is computationally efficient, further optimization through pruning or distillation could enhance performance on ultra-low-power devices. Furthermore, the persistence of some misclassifications highlights the limitation of relying solely on spatial image features. Future research should explore the integration of patient-level clinical metadata, such as age, lesion evolution history, and anatomic site, which dermatologists routinely synthesize to disambiguate visually deceptive lesions. Finally, although we performed external validation on the ISIC 2019 dataset with rigorous deduplication, further multi-center studies on diverse skin phototypes are necessary to confirm global generalizability.

6. Conclusion

This study introduced a lightweight yet robust framework for skin-lesion classification that combined preprocessing, data augmentation, hierarchical two-stage learning, Shuffle Attention and a lesion-focused self-supervised strategy. These components worked together to enhance feature representations, address class imbalance and reduce reliance on large annotated datasets. The proposed model achieved a Top-1 accuracy of 0.961 and an OvO AUC of 0.999 on the HAM10000 dataset, matching or exceeding the performance of state-of-the-art approaches without using external data. Robustness experiments with simulated low-quality images further

showed that the model retained strong discriminative capability under realistic degradation, supporting its suitability for mobile and teledermatology applications.

Author Contributions

C.-H.P. designed and implemented the deep learning models, conducted the experiments and performed the data analysis. T.-K.L. contributed domain expertise in medicine and clinical application, assisted in the interpretation of the experimental results and secured the research funding that supported this work. D.-C.C. conceived and supervised the study, provided domain expertise in medical imaging and image processing, and critically reviewed and edited the manuscript. All authors read and approved the final version of the manuscript. T.-K.L. and D.-C.C. contributed equally.

Competing interests

The authors declare no competing interests.

Data availability

This study used the publicly available HAM10000 dermoscopic image dataset, which was originally published by Tschandl et al. in *Scientific Data* [9]. All data analysed in this work are contained within that dataset. Any additional information is available from the corresponding author upon reasonable request.

Code availability

The code used to implement the proposed YOLOv11-based framework and DS-MAE self-supervised pretraining is openly available at: https://github.com/hank52052/DS-MAE-YOLOv11-Dermoscopy

Funding

This work was supported by the Tainan Municipal An-Nan Hospital, Tainan, Taiwan (grant number ANHRF112-07).

Ethics approval and consent to participate

This study did not involve new data collection from human participants. It was conducted using the de-identified, publicly available HAM10000 dataset originally published by Tschandl et al. [9]. This study was approved by the Institutional Review Board of Tainan Municipal An-Nan Hospital (IRB number TMANH112-REC031).

Supplementary Information

Supplementary material accompanies this paper and includes additional figures and tables that provide extended experimental results and analysis.

References

1. Zhou, L., Zhong, Y., Han, L., Xie, Y. & Wan, M. Global, regional and national trends in the burden of melanoma and non-melanoma skin cancer: insights from the Global Burden of Disease study 1990–2021. Sci. Rep. 15, 5996 (2025).
2. Hu, W., Fang, L., Ni, R., Zhang, H. & Pan, G. Changing trends in the disease burden of non-melanoma skin cancer globally from 1990 to 2019 and its predicted level in 25 years. BMC Cancer 22, 836 (2022).
3. Pega, F. et al. Global, regional and national burdens of non-melanoma skin cancer attributable to occupational exposure to solar ultraviolet radiation for 183 countries, 2000–2019: a systematic analysis from the WHO/ILO Joint Estimates of the Work-related Burden of Disease and Injury. Environ. Int. 181, 108226 (2023).
4. Huang, J. et al. Global incidence, mortality, risk factors and trends of melanoma: a systematic analysis of registries. Am. J. Clin. Dermatol. 24, 965–975 (2023).
5. Rembielak, A. & Ajithkumar, T. Non-melanoma skin cancer–an underestimated global health threat? Clin. Oncol. 31, 735–737 (2019).
6. Trager, M. H., Geskin, L. J., Samie, F. H. & Liu, L. Biomarkers in melanoma and non-melanoma skin cancer prevention and risk stratification. Exp. Dermatol. 31, 4–12 (2022).
7. Esteva, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 542, 115–118 (2017).
8. Mazzetto, R., Sernicola, A., Tartaglia, J., Ciolfi, C. & Alaibac, M. Potential of automated image analysis for the measurement of vitiligo lesions. Front. Med. 12, 1623408 (2025).
9. Tschandl, P., Rosendahl, C. & Kittler, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci. Data 5, 180161 (2018).
10. Brinker, T. J. et al. Deep neural networks are superior to dermatologists in melanoma image classification. Eur. J. Cancer 119, 11–17 (2019).
11. Nahm, W. J., Sohail, N., Burshtein, J., Goldust, M. & Tsoukas, M. Artificial intelligence in dermatology: a comprehensive review of approved applications, clinical implementation and future directions. Int. J. Dermatol. 64, 1568–1583 (2025).
12. Haenssle, H. A. et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann. Oncol. 29, 1836–1842 (2018).
13. Codella, N. et al. Skin lesion analysis toward melanoma detection 2018: a challenge hosted by the International Skin Imaging Collaboration (ISIC). arXiv preprint arXiv:1902.03368 (2019).

14. Combalia, M. et al. Validation of artificial intelligence prediction models for skin cancer diagnosis using dermoscopy images: the 2019 International Skin Imaging Collaboration Grand Challenge. Lancet Digit. Health 4, e330–e339 (2022).
15. Gutman, D. et al. Skin lesion analysis toward melanoma detection: a challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC). arXiv preprint arXiv:1605.01397 (2016).
16. Marchetti, M. A. et al. Results of the 2016 International Skin Imaging Collaboration International Symposium on Biomedical Imaging challenge: comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images. J. Am. Acad. Dermatol. 78, 270–277.e1 (2018).
17. Codella, N. C. F. et al. Skin lesion analysis toward melanoma detection: a challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), hosted by the International Skin Imaging Collaboration (ISIC). In 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) 168–172 (IEEE, 2018).
18. Bakkouri, I. & Afdel, K. MLCA2F: multi-level context attentional feature fusion for COVID-19 lesion segmentation from CT scans. Signal Image Video Process. 17, 1181–1188 (2023).
19. Bakkouri, I., Afdel, K., Benois-Pineau, J. & Catheline, G. BG-3DM2F: bidirectional gated 3D multi-scale feature fusion for Alzheimer's disease diagnosis. Multimed. Tools Appl. 81, 10743–10776 (2022).
20. Bakkouri, I. & Afdel, K. Computer-aided diagnosis (CAD) system based on multi-layer feature fusion network for skin lesion recognition in dermoscopy images. Multimed. Tools Appl. 79, 20483–20518 (2020).
21. Bakkouri, I. & Afdel, K. Multi-scale CNN based on region proposals for efficient breast abnormality recognition. Multimed. Tools Appl. 78, 12939–12960 (2019).
22. Tschandl, P. et al. Human–computer collaboration for skin cancer recognition. Nat. Med. 26, 1229–1234 (2020).
23. Shaheen, H., Singh, M. P. & Singh, A. K. Advanced skin lesion detection via EfficientNetB0 and vision transformer model with spatial-aware attention. Multimed. Tools Appl. 84, 38853–38871 (2025).
24. Ramya, J. & Anil Kumar, K. M. Hybrid approach for skin lesion analysis: integrating modified U-Net segmentation with vision transformers for multi-class skin cancer detection. Int. J. Biomed. Eng. Technol. 48, 337–370 (2025).
25. Manzoor, K., Gilal, N. U., Agus, M. & Schneider, J. Dual-stage segmentation and classification framework for skin lesion analysis using deep neural network. Digit. Health 11, 20552076251351858 (2025).
26. Alghoraibi, H. et al. Deep learning-based Mpox skin lesion detection and real-time monitoring in a smart healthcare system. Diagnostics 15, 2505 (2025).
27. Asif, S., Qurrat-ul-Ain, Khan, S. U. R., Amjad, K. & Awais, M. SKINC-NET: an efficient

693 lightweight deep learning model for multiclass skin lesion classification in dermoscopic
694 images. Multimed. Tools Appl. 84, 12531–12557 (2025).
695 28. Shams, M. Y. et al. Skin disease classification: a comparison of ResNet50, MobileNet and
696 Efficient-B0. J. Curr. Multidiscip. Res. 1, 1–7 (2025).
697 29. Lin, T.-L. et al. Assessing the efficacy of the spectrum-aided vision enhancer (SAVE) to detect
698 acral lentiginous melanoma, melanoma in situ, nodular melanoma and superficial spreading
699 melanoma. Diagnostics 14, 1672 (2024).
700 30. Nazari, S. & Garcia, R. Going smaller: attention-based models for automated melanoma
701 diagnosis. Comput. Biol. Med. 185, 109492 (2025).
702 31. Da Costa Nascimento, J. J. et al. Health of Things Melanoma Detection System—detection
703 and segmentation of melanoma in dermoscopic images applied to edge computing using deep
704 learning and fine-tuning models. Front. Commun. Netw. 5, 1376191 (2024).
705 32. Pirahandeh, M. Dermatological health: a high-performance, embedded and distributed system
706 for real-time facial skin problem detection. Electronics 14, 1319 (2025).
707 33. Hosny, M., Elgendy, I. A., Chelloug, S. A. & Albashrawi, M. A. Attention-based convolutional
708 neural network model for skin cancer classification. IEEE Access 13, 172027–172050 (2025).
709 34. Matiray, S. & Singh, L. K. Deep learning for early skin cancer detection: a comparative study
710 on hybrid CNN models. In Data, Information and Computing Science 111–125 (IOS Press,
711 2025).
712 35. Marie, M. I., Elredeny, M. S. & Yakoub, A. E. A hybrid model for early melanoma detection:
713 integrating YOLOv9 and Faster R-CNN for enhanced diagnostic accuracy. IEEE Access 13,
714 120327–120344 (2025).
715 36. Ozdemir, B. & Pacal, I. A robust deep learning framework for multiclass skin cancer
716 classification. Sci. Rep. 15, 4938 (2025).
717 37. Wen, C. & Abisado, M. B. Modeling skin cancer detection and analysis using improved
718 YOLOv5 towards early diagnosis and clinical applications. In 2025 6th International
719 Conference on Computer Vision, Image and Deep Learning (CVIDL) 279–283 (IEEE, 2025).
720 38. Han, X.-H. & Chen, Y.-W. Decision support system for skin lesion diagnosis using deep
721 learning. In Advances in Intelligent Disease Diagnosis and Treatment: Research Papers in
722 Honour of Prof. Janusz Kacprzyk for Invaluable Contributions 279–304 (Springer, 2024).
723 39. Riaz, S., Naeem, A., Malik, H., Naqvi, R. A. & Loh, W.-K. Federated and transfer learning
724 methods for the classification of melanoma and nonmelanoma skin cancers: a prospective
725 study. Sensors 23, 8457 (2023).
726 40. Patil, A., Mehto, A. & Nalband, S. Enhancing skin lesion diagnosis with data augmentation
727 techniques: a review of the state-of-the-art. Multimed. Tools Appl. 84, 25325–25364 (2025).
728 41. Yadav, R. & Bhat, A. A systematic literature survey on skin disease detection and classification
729 using machine learning and deep learning. Multimed. Tools Appl. 83, 78093–78124 (2024).
730 42. Cassidy, B., Kendrick, C., Brodzicki, A., Jaworek-Korjakowska, J. & Yap, M. H. Analysis of

731 the ISIC image datasets: usage, benchmarks and recommendations. Med. Image Anal. 75,
732 102305 (2022).
733 43. Khan, A. H., Awang Iskandar, D. N. F., Al-Asad, J. F. & El-Nakla, S. Classification of skin
734 lesion with hair and artifacts removal using black-hat morphology and total variation. Int. J.
735 Comput. Digit. Syst. 10, 597–604 (2021).
736 44. Li, Y., Qi, F. & Wan, Y. Improvements on bicubic image interpolation. In 2019 IEEE 4th
737 Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)
738 1316–1320 (IEEE, 2019).
739 45. Buslaev, A. et al. Albumentations: fast and flexible image augmentations. Information 11, 125
740 (2020).
741 46. Khanam, R. & Hussain, M. YOLOv11: an overview of the key architectural enhancements.
742 arXiv preprint arXiv:2410.17725 (2024).
743 47. Zeiler, M. D. ADADELTA: an adaptive learning rate method. arXiv preprint arXiv:1212.5701
744 (2012).
745 48. Smith, L. N. Cyclical learning rates for training neural networks. In 2017 IEEE Winter
746 Conference on Applications of Computer Vision (WACV) 464–472 (IEEE, 2017).
747 49. Zhang, Q.-L. & Yang, Y.-B. SA-Net: shuffle attention for deep convolutional neural networks.
748 In ICASSP 2021 – 2021 IEEE International Conference on Acoustics, Speech and Signal
749 Processing (ICASSP) 2235–2239 (IEEE, 2021).
750 50. Chen, X. et al. Symbolic discovery of optimization algorithms. arXiv preprint
751 arXiv:2302.06675 (2023).
752 51. Paris, S., Kornprobst, P., Tumblin, J. & Durand, F. Bilateral filtering: theory and applications.
753 Found. Trends Comput. Graph. Vis. 4, 1–73 (2009).
754 52. Lan, Z., Cai, S., He, X. & Wen, X. FixCaps: an improved capsules network for diagnosis of
755 skin cancer. IEEE Access 10, 76261–76267 (2022).
756 53. Datta, S. K., Shaikh, M. A., Srihari, S. N. & Gao, M. Soft attention improves skin cancer
757 classification performance. In International Workshop on Interpretability of Machine
758 Intelligence in Medical Image Computing 13–23 (Springer, 2021).
759 54. Gessert, N., Nielsen, M., Shaikh, M., Werner, R. & Schlaefer, A. Skin lesion classification
760 using ensembles of multi-resolution EfficientNets with meta data. MethodsX 7, 100864
761 (2020).
762 55. Halder, A. et al. A fuzzy rank-based deep ensemble methodology for multi-class skin cancer
763 classification. Sci. Rep. 15, 6268 (2025).
764 56. Ji, Z. et al. EFAM-Net: a multi-class skin lesion classification model utilizing enhanced feature
765 fusion and attention mechanisms. IEEE Access 12, 143029–143041 (2024).
766 57. Gururaj, H. L., Manju, N., Nagarjun, A., Manjunath Aradhya, V. N. & Flammini, F. DeepSkin:
767 a deep learning approach for skin cancer classification. IEEE Access 11, 50205–50214 (2023).
768 58. Saha, U. et al. YOLOv8-based deep learning approach for real-time skin lesion classification

using the HAM10000 dataset. In 2024 IEEE International Conference on E-health Networking, Application & Services (HealthCom) 1–4 (IEEE, 2024).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation0428rev.docx