Science Journal of University of Zakho
DOI: https://doi.org/10.25271/sjuoz

Original Article

DEEP LEARNING-BASED SKIN DISEASE DETECTION AND CLASSIFICATION

Zhehat Rebar Abdulqader [1, 2, 3, *], and Araz Rajab Abrahim[4]

[1]Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq.
[2]Department of Information Technology, Technical College of Informatics, Akre University for Applied Science, Akre, Kurdistan Region, Iraq.
[3]College of Education, University of Zakho, Zakho, Kurdistan Region, Iraq.
[4]Dohuk Technical Institute, Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq.

*Corresponding author, Email: zhehat.abdulqader@dpu.edu.krd (Tel: +964-750 780 8945)

Received: 27, Jul, 2025
Accepted: 31, Aug, 2025
Published: 13, Apr, 2026

ABSTRACT

Automatic classification of dermoscopy images is essential for the early diagnosis and treatment of skin diseases. However, this task is challenging due to visual similarities between disease types, variations in skin structures, and differences across disease stages. To address these difficulties, Deep Learning (DL) has emerged as a powerful approach for computer-aided dermatological diagnosis. In this study, we propose a DL framework specifically designed for skin disease classification. The model employs a lightweight ConvNeXt-Tiny architecture, combined with a two-phase hybrid data augmentation strategy and an advanced optimization pipeline. The methodology includes extensive preprocessing of dermoscopic images, followed by hybrid augmentation that merges offline transformations (spatial, pixel-level, structural) with online probabilistic methods such as MixUp and CutMix. This approach improves minority class representation and stabilizes decision boundaries. Experiments on the HAM10000 dataset show strong results: 95.21% accuracy, 93.89% precision, 89.87% recall, 98.62% specificity, 91.60% F1-score, and 98.98% AUC. These outcomes surpass baseline ConvNeXt variants and other state-of-the-art methods. The proposed framework offers a practical solution for deployment in resource-constrained clinical environments, supporting accurate and early diagnosis of skin diseases.

KEYWORDS: CNN, Deep Learning, Skin Disease, ConvNeXt-Tiny, Hybrid Augmentation, Dermoscopy.

1. INTRODUCTION

The human skin, covering approximately 20 square feet, is the largest organ in the body. It serves as a vital protective barrier against disease and plays an important role in regulating body temperature (Karthik *et al.,* 2022). Despite its protective functions, the skin is susceptible to a wide range of disorders due to its complex structure, composed primarily of the epidermis and dermis. Globally, skin diseases are among the most common medical conditions, affecting roughly one-third of the population at some stage in their lives and significantly impacting quality of life (Behara *et al.,* 2023). Excessive exposure to ultraviolet (UV) radiation from the sun is a primary cause of many skin diseases (Kalaiarasan *et al.,* 2022). These conditions, often characterized by uncontrolled proliferation of skin cells, contribute substantially to global morbidity and mortality. Therefore, any cutaneous lesion exhibiting atypical features, the appearance of new growths, or progressive changes in size, shape, or pigmentation should be evaluated by a qualified clinician (Anand *et al.,* 2022). Importantly, early detection of skin diseases can dramatically improve outcomes. For instance, five-year survival rates can increase from less than 14% to nearly 97%, highlighting the critical importance of timely diagnosis (Niino & Matsuda, 2021).

Despite accessibility to images of skin lesions, the traditional diagnostic approach faces notable limitations. Current practices rely heavily on visual inspection and dermoscopic imaging, followed by invasive procedures such as biopsies, interpreted by human experts. These methods are time-consuming, resource-intensive, and inherently subjective, prone to variability depending on the expertise and experience of the clinician (Malik *et al*., 2024). However, even with dermoscopic imaging, accurate classification of skin lesions remains challenging. The global shortage of dermatologists exacerbates the challenge, especially in rural or under-resourced regions, leaving many individuals without timely access to specialized care (Lilhore *et al.*, 2025).

Dermoscopy images are crucial in demonstrating the subtleties in skin lesion diagnosis. Though they are more detailed than standard clinical photos, they still need expert interpretation

Access this article online
https://doi.org/10.25271/sjuoz.2026.14.2.1737
Printed ISSN 2663-628X;
Electronic ISSN 2663-6298
Science Journal of University of Zakho
Vol. 14, No. 02, pp. 370 –385April-2026

and are subject to confounding factors: skin artefacts (hair, bubbles and pigment network), uneven illumination and irregular lesion shapes. To overcome these limitations, researchers are increasingly using the assistance of computer-aided detection (CAD) systems and artificial intelligence (AI)-based approaches to augment clinical diagnosis (Angurana *et al.,* 2019). Traditional machine learning-based approaches were also applied for lesion classification, and these typically utilize hand-designed features based on shape, color, or texture with the sensitivity to noise, background clutter, and variation in lesion appearance (Aboulmira *et al.*, 2024). Recent advancements in Deep Learning (DL) have transformed the domain of medical image analysis. Especially Convolutional Neural Networks (CNNs), which are a type of DL model, take raw image data as input and automatically extract feature representations, avoiding the requirement for manual feature engineering (Kavitha *et al.,* 2024). CNNs have surpassed performance benchmarks for manual processing in imaging, including dermatology, with some models achieving or exceeding the accuracy levels of expert practitioners (Tschandl, 2021). Moreover, it is more efficient and stable to process massive image sets by DL algorithms, which offer faster and more consistent processing (Li *et al.,* 2022). Modern architectures notably GoogleNet, ResNet, Inception, and DenseNet, are frequently employed, pretrained on large natural image corpora and fine-tuned on medical datasets such as HAM10000 (Almutairi & Khan, 2023).

Regardless of the potential that DL offers, certain challenges remain in medical imaging. Perhaps one of the biggest challenges is the small size of medical datasets. For example, the subset of dermatoscopic images often falls short when compared to a dataset like ImageNet (Shahzad *et al.,* 2024). Furthermore, class imbalance can lead to models that are too biased to identify rare, yet critical, lesions (Su *et al.,* 2024). Some of the image artefacts include hairs, air bubbles, and blood vessels, which may obscure crucial components or impede categorization (Hu *et al.,* 2024). In addition, the requirement for models encompasses the phenomena of intra-class variability (visual differences within the same disease category) and inter-class similarity (dissimilar diseases that appear similar), which is also highly demanding (El-fattah *et al.,* 2023). In order to overcome these challenges, DL techniques have been particularly effective with pre-trained CNNs using Transfer Learning (TL) for automated skin lesion classification.

This research aims at fulfilling this immediate need of efficient automated classification of skin lesions by introducing a DL framework. To address the challenges of clinical deployment, our study specifically focuses on a lightweight deep learning model, aiming to achieve a balance between high diagnostic accuracy and computational efficiency. Our framework incorporates the lightweight ConvNeXt-Tiny architecture along with a two-phase hybrid data augmentation method. Our contributions include comprehensive offline augmentations (spatial, pixel-level, structural) to bolster minority classes, followed by online probabilistic MixUp and CutMix during training to regularize decision boundaries during training. The ConvNeXt-Tiny backbone, fine-tuned with AdamW and a OneCycleLR scheduler under mixed-precision arithmetic, has demonstrated superior performance compared to baseline ConvNeXt variants and other modern approaches. The framework's strong generalization capabilities, robustness against class imbalance and image artefacts, and potential for deployment in resource-constrained clinical settings are highlighted through comprehensive evaluation metrics.

The following are the study's main contributions to the classification of skin diseases:

- A tailored strategy was developed for dermoscopic images, combining offline spatial, pixel-level, and structural transforms with probabilistic online MixUp and CutMix augmentations technique. This approach enhanced data diversity and reduced overfitting.

- A lightweight ConvNeXt-Tiny model was fine-tuned using advanced optimization techniques, achieving high accuracy with fast convergence.

- Consistent performance was demonstrated across all lesion categories in the HAM10000 dataset, including rare types.

This paper's remaining sections are organised as follows: Section 2 provides a literature survey of skin disease detection and classification. Materials and the proposed methods are described in detail in Section 3. The results are shown in Section 4. Additionally, Section 5 presents discussion, limitations and future directions. Finally, the main points of conclusion are given in Section 6.

RELATED WORK

Deep learning, particularly CNNs, has significantly advanced dermatological image analysis, often achieving expert-level accuracy. Researchers have developed and adapted sophisticated CNN architectures such as GoogleNet, ResNet, Inception, and DenseNet; they are usually fine-tuned on medical photos for skin lesion classification after being pre-trained on extensive natural image datasets. Building on these advancements, recent studies continue to explore and refine DL approaches to further enhance the performance and reliability of automated skin lesion diagnosis.

For instance, Jain *et al.* (2021) applied six CNN models (VGG19, InceptionV3, ResNet50, Xception, InceptionResNetV2, and MobileNet) with data balancing to HAM10000 dataset, reporting that Xception yielded the best performance. They achieved a top accuracy of about 90.5%, though they observed that the model largely predicted the majority nevus class, indicating bias due to class imbalance. In a similar vein, Alam *et al.* (2022) fine-tuned pretrained models (AlexNet, InceptionV3, and RegNetY-320) on the HAM10000 using image augmentation to address class imbalance. Their best result was obtained with RegNetY-320, reaching 91% accuracy versus 85% for the un-augmented baseline. In contrast, Wei *et al.* (2023) proposed a fusion of DenseNet201 and ConvNeXt-L enhanced with dual attention modules. On HAM10000, their model achieved 95.29% accuracy and an F1-score of 89.99%, outperforming prior baselines. Their study, however, relied on a very large model with nearly 350 million parameters, motivating the development of more lightweight architectures.

Other authors have explored more sophisticated transfer-learning schemes. Ahmad *et al.* (2023) proposed a framework combining data augmentation and two fine-tuned models (Xception and ShuffleNet) on HAM10000. Their ensemble obtained 91.5% accuracy. Alotaibi and AlSaeed (2025) incorporated attention into the Xception backbone: the plain Xception attained 91.05%, while adding self-attention raised accuracy to 94.11%, with soft and hard attention giving 93.29% and 92.97%. Likewise, Ji *et al.* (2024) introduced EFAM-Net, a multi-branch ConvNeXt-based architecture with residual and fusion blocks which reached 93.95% accuracy on HAM10000. These studies demonstrate that transfer learning with attention or augmentation can improve results, yet accuracies remain under 95%.

Vision Transformers (ViTs) have also been applied. Xin *et al.* (2022) introduced "SkinTrans", an enhanced ViT network with multi-scale patch embeddings and contrastive learning. When tested on HAM10000, SkinTrans achieved 94.3% accuracy, markedly higher than earlier CNN baselines. Nevertheless, even this transformer model did not exceed the mid-90s. Complementary to pure-ViT approaches, hybrid

architectures have emerged. Aruk *et al.* (2025) combined ConvNeXt blocks and transformer modules to capture local texture and global context. Their ConvNeXt+ViT hybrid reached 94.30% accuracy on HAM10000. More recently, Liu *et al.* (2025) presented FUSCANet, a lightweight MobileViT-derived network with spatial-channel attention and multi-scale feature aggregation; it achieved 92.80% accuracy on HAM10000 while maintaining a very low parameter count. Despite leveraging modern blocks, these hybrid and transformer-based methods still fall short of the 95% mark.

We observed that despite various advanced DL approaches, including sophisticated CNN architectures, TL with attention mechanisms and data augmentation, and ViTs or hybrid models, a common challenge across these studies was the persistent struggle to achieve accuracy consistently above 95% on the HAM10000 dataset. Furthermore, several works explicitly noted difficulties with minority classes, indicating that class imbalance remained a significant hurdle even when augmentation or attention mechanisms were employed to mitigate it. These challenges inspired us to focus our efforts on addressing this critical issue.

2. MATERIALS AND METHODS:

This section highlights the methodology followed for classifying skin diseases from dermoscopic images, as schematically presented in Figure 1.

Dataset Description:

The HAM10000 dataset was used in this study. It contains 10,015 dermatoscopic images organized into seven different categories of skin lesions. The categories along with their distributions are shown in Table 1, and a sample of images is displayed in Figure 2. The original dermatoscopic images are 600×450 pixels. The images' metadata consist of the diagnoses appertaining to the cases, confirmed either through histopathology or expert review (Tschandl *et al.,* 2018). One of the major challenges with HAM10000, as with many other medical imaging datasets, is severe class imbalance coupled with high interclass similarity. As shown in Table 1, the dataset is highly imbalanced, with common conditions having thousands of images, while rarer conditions have only a few hundred, which can hinder accurate model classification during training.

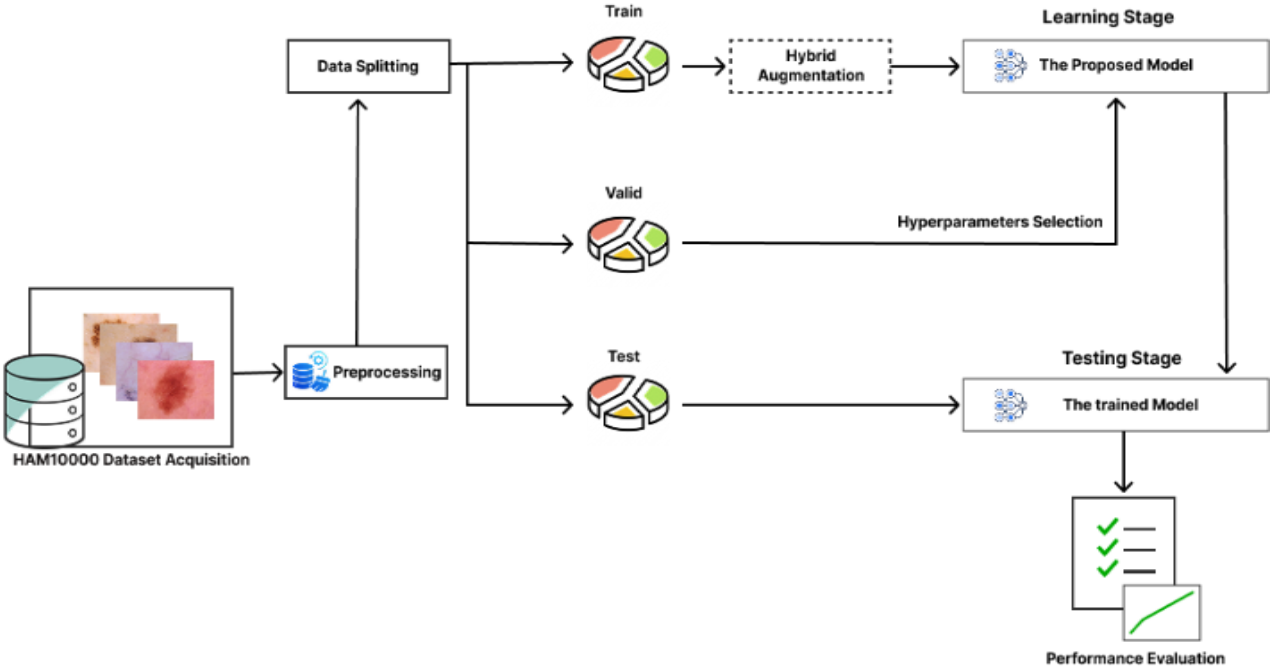

Figure 1: Schematic diagram of the proposed methodology.

Table 1: HAM10000 diseases categories.

Diseases Name	Category Code	No. of Samples	Per. % of Samples
Melanocytic Nevi	nv	6,705	66.95%
Melanoma	mel	1,113	11.11%
Benign Keratosis-Like Lesions	bkl	1,099	10.97%
Basal Cell Carcinoma	bcc	514	5.13%
Actinic Keratoses And Intraepithelial Carcinoma	akiec	327	3.27%
Vascular Lesions	vasc	142	1.42%
Dermatofibroma	df	115	1.15%
Total number of samples		10,015	100.00%

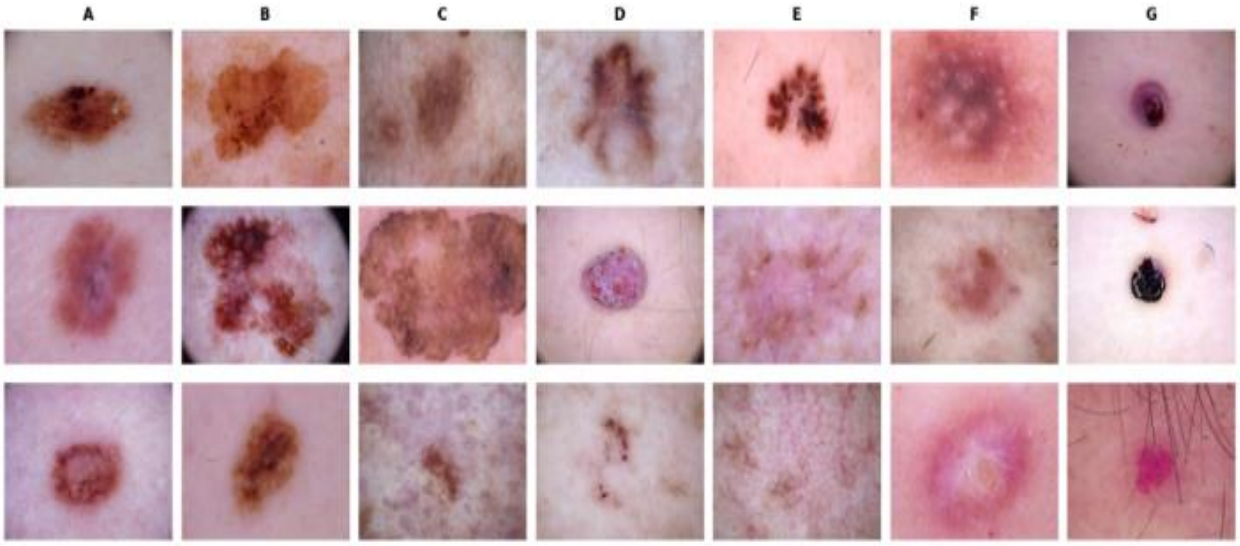

Figure 2: Some sample of HAM10000 dataset: A) Melanocytic Nevi, B) Melanoma, C) Benign Keratosis-Like Lesions, D) Basal Cell Carcinoma, E) Actinic Keratoses, F) Dermatofibroma, and G) Vascular Lesions.

Data Preprocessing and Splitting:

As a first step for model training and to keep input dimensions consistent, all images were randomly cropped down to 384×384. This would ensure that the model's processing was not overloaded and that sampling was consistent. After resizing, pixel-wise normalisation was applied using the formula in Equation 1. The global mean of the dataset (μ) and standard deviation (σ) were calculated over all the pixels in the training set. For each image, the centring and scaling steps were applied so that the image's mean was subtracted and standard deviation was divided, resulting in z-score normalization which is explained below:

$$\text{Z–score} = \frac{x - \mu}{\sigma} \quad (1)$$

where σ is the standard deviation, μ is the mean, and x is the original feature vector. For DL models to learn more quickly and maintain numerical stability, this normalisation step is essential. Normalized images were then randomly partitioned into three independent subsets: training, validation, and testing, as summarized in Table 2. This stratification ensured that each subset accurately reflected the original class proportions. The dataset was divided into an 80%:10%:10% allocation for training, validation, and testing, respectively.

Table 2: The number of instances used as the training, validation and testing sets.

Type of diseases	Training	Validation	Testing	Total
Nv	5,364	670	671	6,705
Mel	890	111	112	1,113
Bkl	879	110	110	1,099
Bcc	411	51	52	514
Akiec	262	33	32	327
Vasc	114	14	14	142
Df	92	12	11	115
Total	8,012	1,001	1,002	10,015
Per %	80%	10%	10%	100%

Data Augmentation Strategy:

As noted in Section 3.1, the HAM10000 dataset suffers from significant class imbalance, which can impede the effective training of DL models. DL models require large and diverse datasets for optimal training due to their dependence on the number of hidden neurones. By expanding the training dataset's size and diversity, data augmentation alleviates these problems. To mitigate the effects of class imbalance, only the training set was subjected to targeted data augmentation approaches. This approach increased the effective sample size of minority classes, leading to a more balanced distribution. As detailed in Table 3, a suite of techniques, including Spatial Transformations, Pixel-Level Augmentations, and Advanced Structural Augmentations were implemented to generate additional samples for minority classes, thereby enhancing model performance. Figure 3 illustrates samples of applied data to augmentation techniques.

Table 3: Summary of data augmentation techniques, including their parameters and intended purposes.

Type of Augment	Augment Name	Parame	Purpose
Spatial Transformations	Horizontal Flip	p=0.5	Flip image horizontally to increase diversity
	Vertical Flip	p=0.5	Flip image vertically for more orientation variance
	Shift Scale Rotate	shift=0.1, scale=0.2, rotate=45, p=0.9	Apply affine transformations (shift, zoom, rotate)
Pixel-Level Augment	Random Brightness Contrast	Brightness=0.2, contrast=0.2, p=0.7	Modify brightness and contrast randomly
	Hue Saturation Value	Hue shift=10, satshift=3, valshift=20, p =0.7	Adjust hue, saturation, and value
	Gauss Noise	var=10, 50, p =0.5	Add random Gaussian noise
	Gaussian Blur	blur=3, p=0.5	Apply Gaussian blur
	Motion Blur	blur=3, p=0.5	Simulate motion blur
Advanced Structural Augment	Grid Distortion	p=0.2	Apply nonlinear grid distortion to the structure
	Coarse Dropout	max=8, min=5, p=0.3	Randomly remove rectangular image regions
	MixUp	alpha=1.0, p=0.3	Linearly blend two images and labels to regularize the model.
	CutMix	alpha=1.0, p=0.2	Insert patches from one image into another, mix labels.

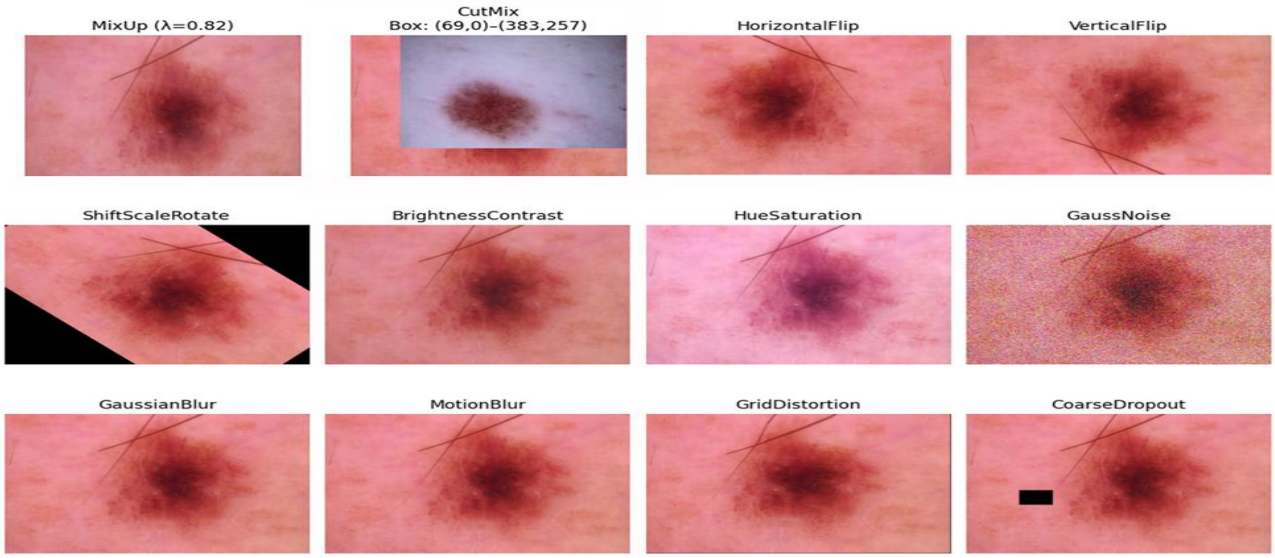

Figure 3: Sample of applied augmentation techniques.

ConvNeXt-tiny:

Our proposed method for skin lesion classification leverages the ConvNeXt-Tiny architecture with the hybrid augmentation, aiming to enhance generalization and performance. The ConvNeXt architecture proposed by Liu *et al.* (2022) is a purely convolutional network designed to bridge the performance gap between convolutional ConvNets and Vision Transformers (ViTs). It integrates ResNet-style residual blocks with innovations inspired by Swin Transformers and self-supervised learning, such as depth-wise convolutions, an inverted bottleneck structure, and refined normalization and activation placement. This design enables ConvNeXt to achieve ViT-like scalability and representational power while retaining the inductive biases of convolutional operations, yielding high accuracy of 87.8% top-1 on ImageNet. ConvNeXt-Tiny, the smallest variant, is chosen for its computational efficiency and suitability for resource-constrained applications like medical image analysis.

The architecture of ConvNeXt-Tiny (Figure 4a) begins with a "stem" module (a 2D convolutional layer followed by Layer Normalization) that processes 384x384x3 input into feature maps. These are then fed through four hierarchical stages, each comprising ConvNeXt blocks and Down-Sample modules. Stages 1 and 2 each contain three ConvNeXt blocks; Stage 3 has nine; and Stage 4 has three. Down-sample modules between stages reduce spatial resolution by 2x and increase channel dimensionality, facilitating abstract representation learning. Following the final stage, a Global Average Pooling (GAP) layer condenses spatial information into a feature vector, which is then passed to a fully connected layer to produce logits for seven skin-lesion categories (nv, mel, bkl, bcc, akiec, vasc, and df).

The core of each stage is the ConvNeXt block (Figure 4b), a modernized ResNet-style bottleneck. It starts with a 7x7 depth-wise 2D convolution for long-range spatial dependencies, followed by Layer Normalization (LN) Equation 2. The features are then projected via two successive 1x1 pointwise

convolutions, separated by a Gaussian Error Linear Unit (GELU) activation Equation 3. A DropPath layer provides stochastic depth regularization, and a residual skip connection maintains gradient flow. The Down-Sample module, Figure 4(c), positioned between stages, uses Layer Normalization followed by a stride-2 pointwise convolution to downscale spatial dimensions and increase channel count, ensuring smooth transitions and maintaining representational capacity between abstraction levels.

$$\text{GELU}(x) = x\,\Phi(x) = x\,\frac{1}{2}\left[1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right] \quad (3)$$

$$\text{LN}(x) = \frac{x-\mu}{\sqrt{\sigma^2+\epsilon}}\,\gamma + \beta \quad (2)$$

where ϵ is the minimum constant for numerical stability, μ and σ are the mean and standard deviation of the input x, and γ and β are the scaling and shifting parameters that are learnt by the model.

Where x is the activation function's input, $\Phi(x)$ is the standard normal distribution's cumulative distribution function (or CDF), and $\text{erf}(x)$ is the Gauss error function, which is associated with the likelihood that a value in a normal distribution is less than x.

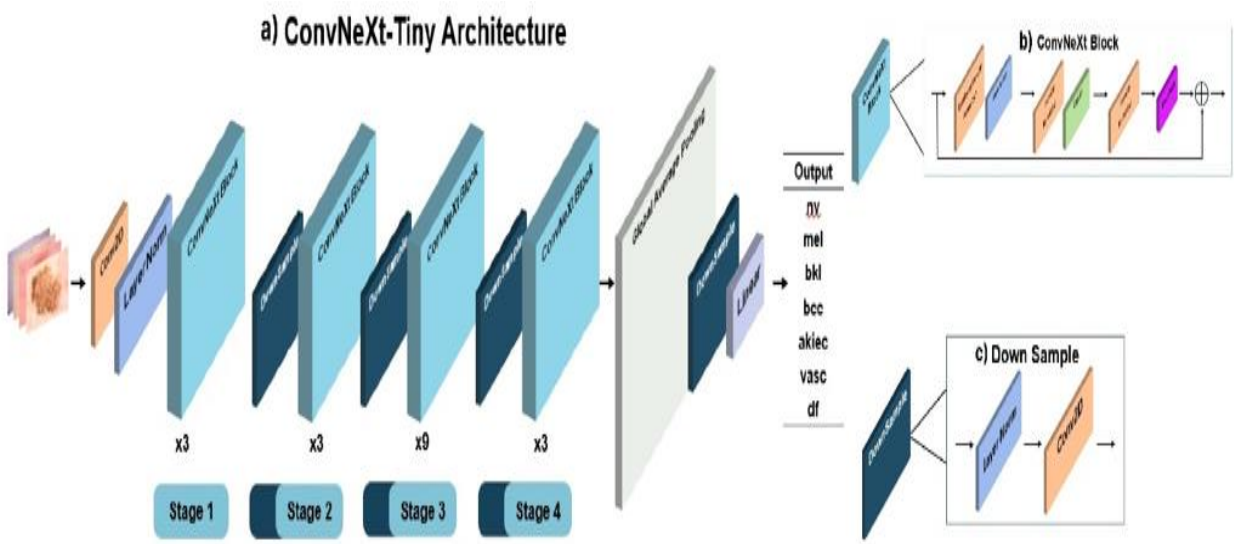

Figure 4: Components and structure of the ConvNeXt-Tiny network.

Proposed Method:

In this study, as depicted in Figure 5, we propose an end-to-end skin lesion classification framework that brings together the ConvNeXt-Tiny network with a two-phase hybrid data augmentation pipeline in pursuit of improved generalization of the network as well as its performance.

The workflow began with careful data preprocessing and splitting, as detailed in Section 3.2. Then, our proposed two-phase hybrid augmentation approach was applied in two consecutive phases. In the first one executed offline, we utilize the Albumentations library in applying a whole set of transformations. This one contains spatial transformations (e.g., flips, rotations, and shifts) as well as pixel-level improvements (e.g., brightness, contrast, and saturation modifications). In addition, advanced structural augmentations, such as grid deformation as well as coarse dropout, are also utilized as means of simulating diverse image perturbations in realistic settings in order to make the model more resilient against irregularities as well as noise. We mention specific detailed types as well as parameters of such augmentation procedures in Section 3.3.

In the second stage, online probabilistic augmentations are applied in a probabilistic manner to each mini-batch at training time. That is, each mini-batch will have a 30% chance of being subject to MixUp augmentation, a 20% chance of being subject to CutMix, and a 50% chance of a routine forward pass without these combined augmentations. By linearly interpolating image pairs and their labels, MixUp generates interpolated examples, expanding the training distribution and promoting linear behaviour among training examples. In contrast, CutMix cuts out patches of one image and pastes them into another image, interpolating their labels proportionally with respect to patch area. This method forces the network to attend to multiple discriminative regions throughout an image, improving its feature recognition of localized features. Both MixUp and CutMix act as effective regularization tools, inducing smoother decision boundaries and generalizing better across models due to decreased overfitting.

We employed the pretrained ConvNeXt-Tiny architecture as our backbone framework, initialized with ImageNet weights to leverage its effectiveness and strong generalization capabilities. AdamW was used as the optimizer, combined with a OneCycleLR learning rate scheduler to ensure effective convergence. To reduce overconfidence and improve model calibration, we adopted a smoothed categorical cross-entropy loss function. In addition, mixed-precision training was applied to accelerate training and reduce memory consumption (see configuration details in the following section). This integrated framework, consisting of a two-phase augmentation strategy and the efficient pretrained ConvNeXt-Tiny backbone, led to significant improvements in skin lesion classification accuracy on the HAM10000 dataset.

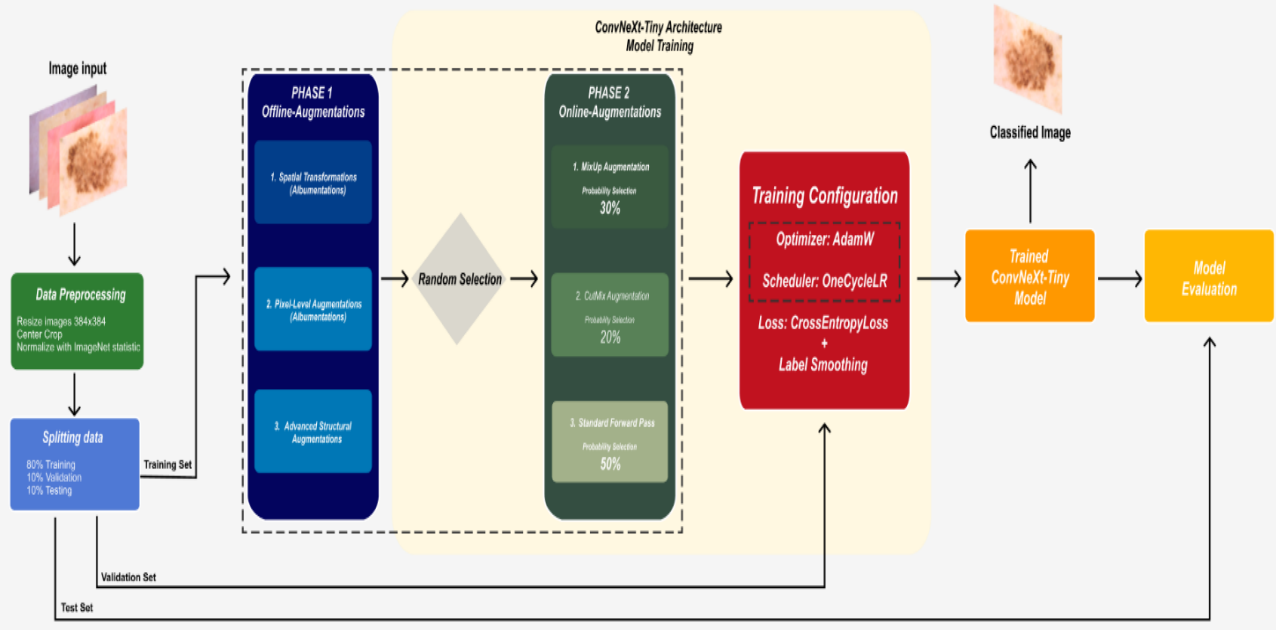

Figure 5: Proposed Architecture Method.

Table 4: The configurations of the ConvNeXt-tiny model.

Params	Amount
Number of blocks	(3, 3, 9, 3)
Number of Channels	(96, 192, 384, 768)
Total params	27,825,511
Trainable params	27,825,511
Non-trainable params	0

Table 5: Architecture of the proposed ConvNeXt-Tiny models for classifying skin diseases.

Layer	Input Shape	Output Shape	Param #
ConvNeXt	[1, 3, 384, 384]	[1, 7]	--
Sequential: 1-1	[1, 3, 384, 384]	[1, 96, 96, 96]	--
Conv2d: 2-1	[1, 3, 384, 384]	[1, 96, 96, 96]	4,704
LayerNorm2d: 2-2	[1, 96, 96, 96]	[1, 96, 96, 96]	192
Sequential: 1-2	[1, 96, 96, 96]	[1, 768, 12, 12]	--
ConvNeXtStage: 2-3	[1, 96, 96, 96]	[1, 96, 96, 96]	237,888
ConvNeXtStage: 2-4	[1, 96, 96, 96]	[1, 192, 48, 48]	992,256
ConvNeXtStage: 2-5	[1, 192, 48, 48]	[1, 384, 24, 24]	11,112,960
ConvNeXtStage: 2-6	[1, 384, 24, 24]	[1, 768, 12, 12]	15,470,592

DOI: https://doi.org/10.25271/sjuoz.2026.14.2.1737

Identity: 1-3	[1, 768, 12, 12]	[1, 768, 12, 12]	--
NormMlpClassifierHead: 1-4	[1, 768, 12, 12]	[1, 7]	--
SelectAdaptivePool2d: 2-7	[1, 768, 12, 12]	[1, 768, 1, 1]	--
LayerNorm2d: 2-8	[1, 768, 1, 1]	[1, 768, 1, 1]	1,536
Flatten: 2-9	[1, 768, 1, 1]	[1, 768]	--
Identity: 2-10	[1, 768]	[1, 768]	--
Dropout: 2-11	[1, 768]	[1, 768]	--
Linear: 2-12	[1, 768]	[1, 7]	5,383

Model Compression and Optimization:

To enhance the suitability of our lightweight ConvNeXt-Tiny model for resource-constrained clinical environments, we implemented two primary model compression techniques: dynamic quantization and unstructured magnitude pruning.

Dynamic quantization is the technique that reduces the model size by converting floating-point weights and activations to a lower-bit integer format (e.g., 32-bit float to 8-bit integer) at runtime. This process significantly decreases the memory footprint of the model, which is crucial for deployment on edge devices with limited storage and memory.

Unstructured magnitude pruning is a method for reducing the number of parameters in a neural network by removing weights with the smallest magnitude. We applied unstructured magnitude pruning with a sparsity ratio of 30%, where individual, low-magnitude weights across the model's layers are permanently set to zero. This results in a sparser model, leading to potential reductions in computational complexity and a decrease in inference time.

We evaluated the performance of each technique individually and in a combined approach (pruning followed by quantization) to assess their impact on model accuracy, size, and computational efficiency.

Training Configuration:

To initiate the setup configuration, the ConvNeXt-Tiny backbone is used to extract robust features, specifically after the last convolutional block, where the standard classification head is substituted. At this point, a Global Average Pooling (GAP) layer is applied to the feature maps. This process computes the average of each feature map, which drastically decreases the complexity of the feature vectors while keeping critical spatial information. The pooled features are then sent straight into a SoftMax classifier, which generates a probability distribution across the various skin lesion types, allowing for exact disease classification. Inference time was measured to evaluate the computational performance of the trained model, using a Kaggle P100 GPU and a batch size of 64 on the entire test dataset. For a detailed review of the hyperparameters and their corresponding configurations used in training our proposed model, refer to Table 6.

Furthermore, a smoothed categorical cross-entropy loss presented in Equation 4, which integrates standard cross-entropy with label smoothing, was employed to replace hard one-hot targets with softened distributions. The original loss for a single example with one-hot label $y = (y_1,\ldots, y_k)$ and prediction probability $p = (p_1,\ldots, p_k)$ is given by $\text{Loss}_{CE}(y, p) = -\sum_{i=1}^{k} y_i \log p_i$ with label smoothing of factor ε the target becomes $\tilde{y}_i = y_i(1-\varepsilon) + \frac{\varepsilon}{K}, i = 1, \ldots, k$, and the smoothed loss is:

$$\text{Loss}_{CE}(\tilde{y}, p) = -\sum_{i=1}^{k}\left[y_i(1-\varepsilon) + \frac{\varepsilon}{K}\right] \log p_i \qquad (4)$$

In our implementation (ε=0.1), 90% of the probability mass remains on the true class, and 10% is uniformly spread across the others. This regularization reduces overconfidence, mitigates class-imbalance effects, yields smoother loss curves, and accelerates convergence, resulting in improved validation accuracy compared to the conventional formulation. Label smoothing also produces better-calibrated probabilities, essential for clinical decision support. By softly redistributing a small fraction of probability mass, label smoothing helps counteract the pronounced class imbalance in the dataset, ensuring under-represented categories receive sufficient gradient signal.

Table. 6: Hyperparameters used during training.

Hyperparameters	Value
input size	384 × 384
Scheduler	OneCycleLR
Learning rate	5×10^{-5}
Batch size	64
Momentum (β_1)	0.9

DOI: https://doi.org/10.25271/sjuoz.2026.14.2.1737

Optimizer	AdamW
Epochs	100
Weight decay	5×10^{-3}
Warmup epochs	30
Warmup learning rate	5×10^{-6}
Loss function	categorical cross-entropy + label smoothing (0.1)

Performance Evaluation Metric:

To evaluate the performance of the proposed dermoscopic skin lesion classification model in a multiclass setting, we employ several standard metrics derived from the confusion matrix, including overall accuracy, precision, recall (sensitivity), specificity, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC). Overall accuracy measures the proportion of correctly classified cases (Equation 5). Precision quantifies the proportion of true positive predictions among all positive predictions (Equation 6), while recall, or sensitivity, reflects the model's ability to identify actual positive cases (Equation 7). Specificity assesses the proportion of true negatives correctly identified (Equation 8), and the F1-score provides a balanced measure of precision and recall through their harmonic mean (Equation 9). Finally, AUC evaluates classification performance across the full range of threshold values (Equation 10). Together, these metrics provide a comprehensive assessment of both per-class performance and the overall discriminative capability of the model.

$$\text{Accuracy} = \frac{\text{TP+ TN}}{\text{TP+TN+FP+FN}} \tag{5}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+ FP}} \tag{6}$$

$$\text{Recall} = \text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \tag{7}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \tag{8}$$

$$\text{F1-Score} = \frac{2 \times \text{Percision} \times \text{Recall}}{\text{Percision + Recall}} \tag{9}$$

$$\text{AUC/ROC} = \int_0^1 TPR(FPR)\, d(FPR) \tag{10}$$

3. RESULTS

This section provides a thorough assessment of the proposed method for skin disease classification, followed by a comparison with current modern methods. Every experiment was performed using an NVIDIA Tesla P100 GPU running Python 3.8 and equipped with 16 GB of dedicated VRAM and 13 GB of host memory on the Kaggle platform. The PyTorch framework, which supports parallel processing with CUDA 11.2, was used to design and train the proposed method.

Overall Performance of the Proposed Models:

More exploration of the entire performance of the resultant model was conducted with widely reported measures of performance, as presented in section 3.8 from equation 5 through equation 10, respectively. These findings can be observed in Table 7. Here, it is evident that the resulting model attains good precision, specificity, AUC, and global accuracy along with good discriminative capacity.

Table 7: Performance analysis of the proposed model under different evaluation metrics.

Metrics	Acc	Prec	Rec	Spec	F1	AUC
Proposed Model	95.21	93.89	89.87	98.62	91.6	98.98

Table 8 shows the performance of the proposed method for the various individual disease types with multiple metrics of evaluation. By providing our evaluation in this manner, we can see how well the compared model, random forest, performs for all of the different conditions and where refinements may be most useful. For example, the model is performing exceptionally well

in NV and VASC having very high accuracy, recall, and F1-scores, but relatively low accuracy, recall, and F1-scores for MEL and AKIEC. Presenting the results in this fashion is an important consideration to understand the diagnostic utility of the model and to evaluate on a more granular level the particular disease outcomes where complete optimization may still be necessary. It is important to note that specificity and AUC appear to be the strongest metrics, while recall was still acceptable at a distinct level below others since the importance of developing accurate results for the less common lesion types was more difficult to perform.

Table. 8: Performance Analysis of Proposed Method for Disease Metrics.

Cate.	Acc	Prec	Rec	Spec	F1	AUC
nv	98.96	96.65	98.96	93.05	97.79	98.24
mel	82.14	89.32	82.14	98.76	85.58	96.71
bkl	90.00	94.29	90.00	99.33	92.09	98.76
bcc	98.08	91.07	98.08	99.47	94.44	99.96
akiec	78.12	92.59	78.12	99.79	84.75	99.27
vasc	98.96	93.33	100.0	99.90	99.90	100.0
df	81.82	100.0	81.82	100.0	90.00	100.0

Furthermore, a comparative analysis of different components of the proposed method, including the base ConvNeXt-Tiny, ConvNeXt-Tiny with offline augmentation, and ConvNeXt-Tiny with online probabilistic augmentation, is presented in Table 9, which demonstrates that our proposed two-phase hybrid augmentation strategy significantly enhances performance. The combination of both offline and online augmentation, as implemented in the full proposed method, achieved the highest performance across all metrics. This incremental improvement highlights the synergistic effect of the two-phase augmentation strategy, where each component contributes to enhancing the model's learning and generalization capabilities.

Table 9: Different ablation of proposed method.

Models	Acc	Prec	Rec	Spec	F1	AUC
Base ConvNeXt-Tiny	92.81	90.73	85.42	98.02	87.86	98.60
ConvNeXt-Tiny + Offline augmentation	93.71	93.91	90.74	92.19	98.05	98.75
ConvNeXt-Tiny + Online probabilistic augmentation	93.61	91.56	89.37	98.25	90.22	98.78
Proposed Method	95.21	93.89	89.87	98.62	91.60	98.98

Training and Validation Performance Analysis:

Figure 6 presents a comparative analysis of the training and validation performance across four distinct ablations of the proposed method with ConvNeXt-Tiny. The plots consistently show that all models exhibit a decreasing trend in both training and validation loss, alongside an increasing trend in accuracy over epochs. However, significant differences in convergence and generalization are observable. As depicted, (a) demonstrates the most stable and highest validation accuracy, consistently surpassing the training accuracy and reaching a peak around 0.95. This is because data augmentation techniques such as CutMix and MixUp were applied only to the training data, enhancing its diversity and making the training task slightly harder, while the validation set remained unaltered. The model also maintains a low and converging validation loss, indicating its ability to learn representative features rather than merely memorizing the training data. This reflects superior generalization and effective mitigation of overfitting. In contrast, (c) shows a clear gap between high training accuracy and significantly lower validation accuracy, coupled with a higher and more volatile validation loss, suggesting a tendency towards overfitting without robust augmentation. Both (b) and (d) show improvements over the base model, with (b) demonstrating slightly better stability in validation loss and accuracy compared to (d). The superior performance of the proposed method underscores the synergistic benefits of combining both offline and online augmentation strategies, leading to a more robust and generalizable model.

DOI: https://doi.org/10.25271/sjuoz.2026.14.2.1737

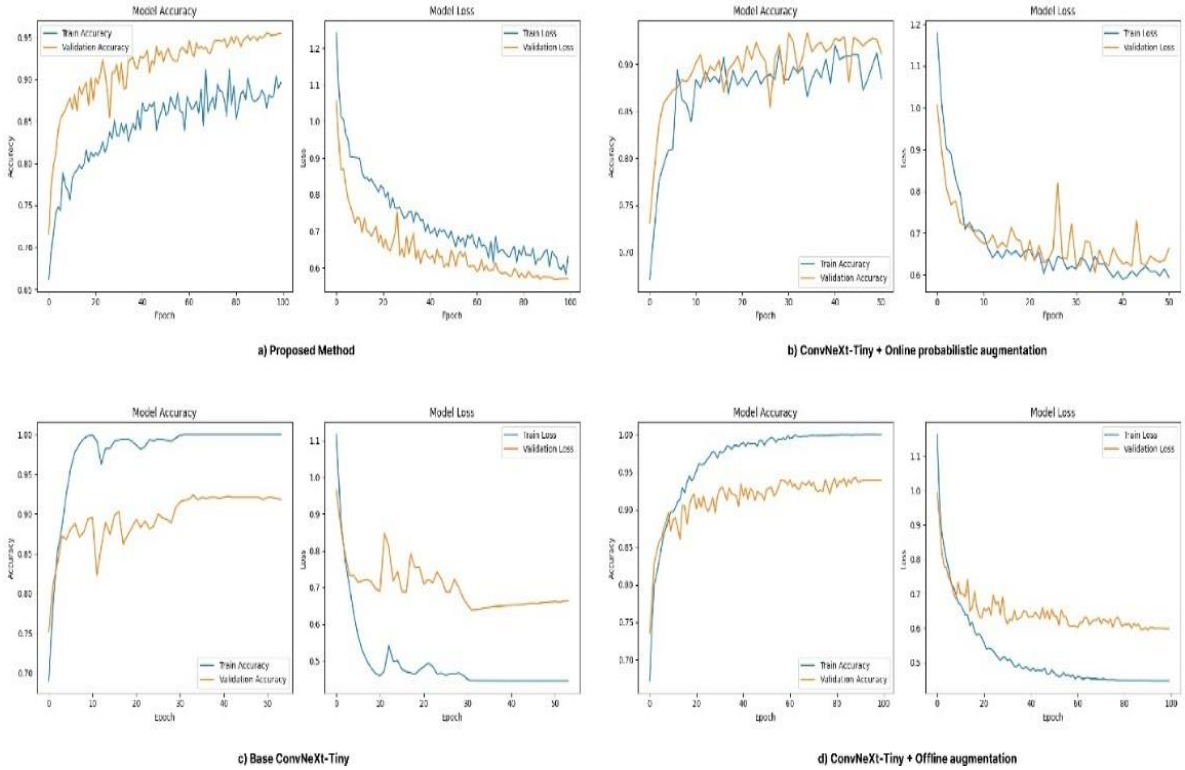

Figure 6: Comparative training and validation performance of ConvNeXt-tiny with various data augmentation strategies.

Model Interpretability Using Grad-CAM:

To ensure transparency and reliability of the proposed model's decision-making process, we incorporated an interpretability step based on Gradient-weighted Class Activation Mapping (Grad-CAM). Grad-CAM is a widely used post-hoc visualization technique that highlights the discriminative regions of an input image that contribute most to the model's prediction. It operates by computing the gradient of the target class score with respect to the feature maps of the final convolutional layers. These gradients are then globally averaged to obtain weights, which are combined with the corresponding activation maps to generate a class-discriminative heatmap.

In this study, after training the hybrid CNN model on the HAM10000 dataset, Grad-CAM was applied to dermoscopic images from the test set. The resulting heatmaps were further overlaid on the original images to highlight the regions of attention. As illustrated in Figure 7, the model consistently focused on the lesion areas rather than background artefacts such as skin texture or hair. This interpretability step provides confidence that the model's predictions are based on clinically relevant features. Moreover, it enhances trustworthiness, which is crucial for deploying deep learning-based systems in real-world medical applications.

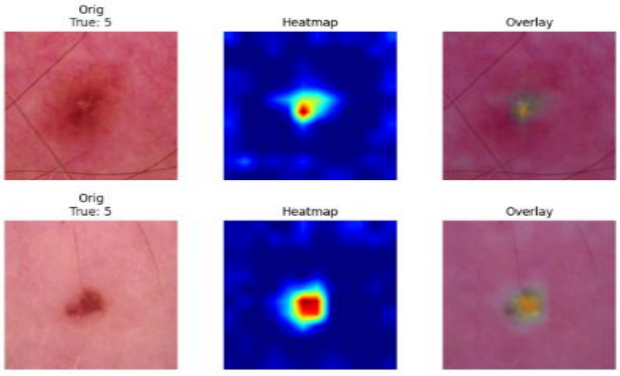

Figure 7: Heatmap Visualizations for Model Interpretability.

Computational Performance:

In addition to the classification performance, an analysis of the model's computational efficiency was conducted to address its suitability for real-world, resource-constrained applications. We measured the inference time of the trained ConvNeXt-Tiny model on a Kaggle P100 GPU. The results are summarized in Table 10, which yielded an avrage inference time per batch (64 test images) of 642.27 ms over 100 runs. This translates to an inference time per image of approximately 10.035 ms. This analysis confirms that the lightweight ConvNeXt-Tiny architecture not only achieves high diagnostic accuracy but also offers valuable computational metrics, making it a viable solution for deployment in clinical setting.

Table 10: ConvNeXt-Tiny Model Inference Performance.

Metric	Value
Avg. inference time per batch	642.27 ms
Avg. inference time per sample	10.035 ms

Performance Analysis of Optimized Models:

We conducted a detailed evaluation of the base, quantized, and pruned models, focusing on accuracy, inference time, model size, and parameter count, as shown in Table 11. All GPU-based experiments were performed on an NVIDIA Tesla P100, while quantized models were evaluated on a CPU, reflecting typical deployment scenarios.

Dynamic quantization demonstrated a substantial 93.1% reduction in model size, decreasing from 106.15 MB to 7.28 MB, and reduced the number of parameters from 27,825,511 to 1,908,480, with only a minor accuracy drop of 0.004. This highlights the effectiveness of quantization for reducing memory footprint, making it highly suitable for deployment on memory-constrained devices or CPUs. However, the inference time of the

DOI: https://doi.org/10.25271/sjuoz.2026.14.2.1737

quantized model was significantly higher than the base model because the dynamic quantization library used does not currently support GPU acceleration.

In contrast, pruning with 30% sparsity resulted in a significant 1.53× speedup in inference time, lowering the average batch processing time from 642.27 ms to 420.90 ms, while slightly improving accuracy by 0.002. Importantly, the model size remained unchanged after pruning. This is because pruning only sets the weights with the least contribution to zero, rather than physically removing them from the model. As a result, pruning primarily benefits GPU inference speed rather than memory reduction.

Overall, the results highlight a clear trade-off between the two optimization techniques. Quantization is ideal for minimizing model size for CPU or edge deployment, though at the cost of slower inference when GPU support is unavailable. Pruning, on the other hand, effectively accelerates GPU inference without reducing memory usage.

Table 11: Comprehensive Comparison of Model Optimization Results.

Model Type	Accuracy	Inference Time (ms/batch)	Device	Model Size (MB)	Parameters
Base	0.9521	642.27	GPU	106.15	27,825,511
Quantized	0.9481	22382.76	CPU	7.28	1,908,480
Pruned	0.9541	420.90	GPU	106.15	27,825,511

Ablation Study on Backbone Variants, Input Resolution, Optimizers, and Batch Size:

For further evidence, we conducted comprehensive ablation studies on various setup settings and model architectures. The investigation was dependent on several key settings, as discussed and presented in section (3.6). We investigated the impact of different ConvNeXt versions, as shown in Table 12, with the base model of ConvNeXt-Tiny consistently yielding the best performance across metrics like Accuracy, Precision, and AUC. Similarly, we explored the influence of different input sizes, as detailed in Table 13, revealing that an input size of 384 achieved the highest scores. We also examined various optimization techniques, presented in Table 14, where AdamW demonstrated superior performance compared to Adam, Nadam, and SGD. Furthermore, Table 15 illustrates the effect of different batch sizes, confirming that a batch size of 64 provided the most effective configuration for the task. These findings reinforced our selection of the Base model ConvNeXt-Tiny with a 384x384 input size, AdamW as the optimizer, and a batch size of 64 as the most effective combination for the proposed method.

Table 12: Proposed method on different ConvNeXt versions.

Models	Acc	Prec	Rec	Spec	F1	AUC
ConvNeXt_Small	91.52	88.62	85.72	97.25	86.66	97.01
ConvNeXt_Base	91.86	89.02	85.93	97.46	87.12	97.22
ConvNeXtV2_ Base	92.51	89.45	87.54	97.84	88.10	97.74
ConvNeXtV2_ Tiny	93.61	91.97	89.54	98.18	90.61	98.81
Base model ConvNeXt-Tiny	95.21	93.89	89.87	98.62	91.60	98.98

Table 13: Proposed method on different input size.

Input Size	Acc	Prec	Rec	Spec	F1	AUC
128	91.32	85.29	85.22	97.62	85.06	97.66
192	92.32	90.51	85.68	97.83	87.90	98.24
256	92.02	89.96	86.12	97.54	87.69	98.24
384	**95.21**	**93.89**	**89.87**	**98.62**	**91.60**	**98.98**

DOI: https://doi.org/10.25271/sjuoz.2026.14.2.1737

Table 14: Proposed method on a different optimization technique.

Optimization	Acc	Prec	Rec	Spec	F1	AUC
Adam	89.97	87.8	81.85	97.06	83.4	98.67
Nadam	88.92	85.41	79.62	97.18	81.66	98.53
SGD	86.73	80.21	70.24	96.29	74.12	97.69
AdamW	95.21	93.89	89.87	98.62	91.6	98.98

Table 15: Proposed method on different batch size.

Batch Size	Acc	Prec	Rec	Spec	F1	AUC
8	93.01	90.35	88.86	98.01	89.42	--
16	93.21	92.59	89.64	97.96	91.00	--
32	94.21	90.33	90.06	98.55	90.07	--
64	95.21	93.89	89.87	98.62	91.60	98.98

4. DISCUSSION

The primary focus of this section is to conduct a thorough investigation of the competitive performance of our hybrid ConvNeXt Tiny framework that we trained on the HAM10000 dataset, considering its place among other transfer learning strategies competing with it, which includes both general-purpose CNNs and specialized variants of ConvNeXt, as demonstrated in Tables 14 and 15. It is evident that both specialized and general automated skin disease classifiers have their own unique strengths. The significance of this study goes beyond just comparing the proposed method's performance to other benchmarked solutions; it reveals the importance of design decisions and training strategies utilized.

The results obtained from the HAM10000 dataset can clearly demonstrate the power of our two-phase augmentation methods with ConvNeXt-Tiny architecture to achieve adjustments. For instance, achieving a 95.21% accuracy, 93.89% precision, 89.87% recall, 98.62% specificity, 91.60% F1 score, and 98.98% AUC of the model indicates its extreme strength in discriminating between a wide variety of skin disorders and its ability to generalize across diverse skin conditions.

When comparing our method against established transfer learning architectures in Table 14, all studies included in the comparison also utilized the HAM10000 dataset, ensuring a fair evaluation. EfficientNetV2 stands out with a high-performance record, e.g., 94.0% accuracy as well as 99.3% AUC. With other models, including DenseNet169 as well as RegNetY-320, have also reported good results. In this scenario, however, EfficientNetV2 repeatedly exhibited a superior, well-balanced performance across main measures. Specifically, certain strategies, including CNN inherent learning, despite a good accuracy of 92.89%, exhibited significantly low recall 58.57% as well as an F1-score of 63.14%, exhibiting struggles detecting a significant number of positive instances, a crucial element in medical diagnosis. DenseNet201 and ConvNeXt-L which have superior accuracy of 95.29%; however, the number of parameters in ConvNeXt-L take much time to train. Overall performance of the proposed method in Table 7 exhibits superior performance over most of the models in Table 16 in accuracy as well as F1-score, exhibiting superior competitiveness even against heterogeneous foundation architectures.

The true power of our approach is most evident when seen in relation to other ConvNeXt-based models in Table 15. Our approach is superior to CAFormer-s32 with ConvNeXt block, ConvNeXt-ST-AFF, DenseNet and ConvNeXt Fusion, EFAM-Net, and FUSCANet in almost all reported indicators, including accuracy, precision, recall, F1-score, and AUC. Its excellence can be attributed primarily to such unique architectural breakthroughs as well as special training protocols, in particular, in the two-phase hybrid data augmentation scheme. High precision as well as recall with superior specificity and AUC reveal the potential of the obtained model in precise diagnosis as well as low likelihood of false-positive or false-negative rates, most essential in medical decision-making. The consistent strength of our proposed method, not only against a range of general transfer learning models but especially within the advanced ConvNeXt family, underscores its significant potential for real-world application in clinical settings. The comprehensive evaluation across various metrics demonstrates its balanced diagnostic capabilities, suggesting that it can serve as a reliable automated diagnostic tool for skin diseases, ultimately supporting earlier detection and more effective patient care, particularly in resource-constrained environments.

Despite its superior performance, the current work has some intrinsic drawbacks. First, all experimentation was conducted only using the HAM10000 dataset, thus not allowing generalizing of results. Second, while the problem of class imbalance was alleviated partially using a two-phase hybrid augmentation method, extremely rare classes such as dermatofibroma (1.15% of samples) tended to have lower recall and F1-scores. It is a sign of persistent under-representation bias, with an urgent need to study this issue of imbalanced datasets in due course.

Overcoming such disadvantages, future work will focus on a few central directions. In order to make our solutions not only more resilient but also generalizable, we will perform multi-center testing on diverse dermoscopic collections. For improved generalizability towards little-labelled rare lesions, we will investigate semi-/self-supervised learning avenues. In architecture, our work will examine lightweight transformer-augmented architectures as well as model-compression approaches towards facilitating deployment on point-of-care devices with relative ease, including techniques such as

DOI: https://doi.org/10.25271/sjuoz.2026.14.2.1737

quantization and pruning to further reduce model size while maintaining performance. Finally, towards increased transparency as well as clinician uptake, we will implement intrinsic explainability in our systems with attention-based or concept-bottleneck-based mechanisms.

Table 16: Performance comparison of the existing methods.

Models	Acc	Prec	Rec	Spec	F1	AUC
DenseNet169 (Gururaj *et al.,* 2023)	91.20	--	--	--	91.70	--
RegNetY-320 (Alam *et al.,* 2022)	91	--	--	--	88.1	95
Xception (Kalaiarasan *et al.,* 2022)	90.48	88.76	89.57	--	89.02	--
ResNet50 (Anand *et al.,* 2022)	90	71.28	74.42	--	72.10	--
EfficientNetV2 (Hu *et al.,* 2024)	94.0	91.2	91.7	98.2	91.3	99.3
EfficientNet B5 (Ali *et al.,* 2022)	87.62	88	88	88	87	97.54
CNN inherent learning (Hosny *et al.,* 2024)	92.89	76.85	58.57	95.57	63.14	--
DenseNet201 and ConvNeXt-L (Wei *et al.,* 2023)	95.29	--	--	--	89.99	--

Table 17: Performance comparison of ConvNeXt models.

Models	Acc	Prec	Rec	Spec	F1	AUC
CAFormer-s32 with ConvNeXt block (Aruk *et al.,* 2025)	94.30	93.00	89.59	--	91.11	--
ConvNeXt-ST-AFF (Hao *et al.,* 2023)	92.16	90.96	87.08	98.54	88.83	
DenseNet and ConvNeXt Fusion (Wei *et al.,* 2023)	90.85	83.75	83.81	--	83.45	--
EFAM-Net (Ji *et al.,* 2024)	93.95	90.23	91.44	98.44	90.78	--
FUSCANet (Liu et al., 2025)	92.80	88.43	87.41	97.99	87.78	--
Proposed Method	95.21	93.89	89.87	98.62	91.60	98.98

CONCLUSION

Our study introduces a robust DL framework for accurate and efficient automated skin lesion classification, effectively addressing common challenges in dermoscopic datasets like HAM10000, including limited size, class imbalance, and visual similarities between disease types. Our approach leverages the lightweight ConvNeXt-Tiny architecture combined with a unique two-phase, hybrid data augmentation strategy. This strategy incorporates diverse offline transformations (spatial, pixel-level, structural) to enrich under-represented classes and online probabilistic techniques (MixUp and CutMix) during training to regularize decision boundaries.

The framework's advanced optimization pipeline, featuring the AdamW optimizer, a OneCycleLR scheduler, mixed-precision training, and label-smoothed cross-entropy loss, further enhances performance and prediction calibration while maintaining practical deployability. Comprehensive evaluation using metrics such as accuracy, precision, recall, F1-score, and AUC demonstrates that our framework surpasses baseline ConvNeXt variants and several existing modern methods. The model consistently exhibits balanced and resilient performance across all seven HAM10000 diagnostic categories, highlighting its potential for practical deployment in resource-constrained clinical settings to improve early diagnosis and identification of skin diseases. The successful application of dynamic quantization and magnitude pruning establishes a new benchmark for resource-efficient skin disease classification. This work demonstrates that high-performance deep learning models can be effectively compressed without a significant loss of accuracy, thereby bridging the gap between state-of-the-art research and real-world clinical deployment.

Acknowledgment:

The authors acknowledge all individuals and institutions that supported the completion of this study.

Author Contribution:

Z.R.A., Conceptualization, methodology, software implementation, data curation, formal analysis, visualization, and original draft preparation. A.R.A., Supervision, validation,

investigation, resources, review and editing. All authors have read and agreed to the published version of the manuscript.

Ethical Approval:

This study did not involve any experiments on humans or animals and did not require ethical approval. All data used in this research were obtained from publicly available international datasets and used in accordance with the terms and conditions stated by the dataset providers.

Conflict of Interest:

The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding:

This research did not receive any specific funding from public, commercial, or non-profit organizations.

REFERENCES

Aboulmira, A., Hrimech, H., & Lachgar, M. (2024). Skin Diseases Classification with Machine Learning and Deep Learning Techniques: A Systematic Review. *International Journal of Advanced Computer Science and Applications*, *15*(10), 1155–1173. https://doi.org/10.14569/IJACSA.2024.01510118

Ahmad, N., Shah, J. H., Khan, M. A., Baili, J., Ansari, G. J., Tariq, U., Kim, Y. J., & Cha, J. H. (2023). A novel framework of multiclass skin lesion recognition from dermoscopic images using deep learning and explainable AI. *Frontiers in Oncology*, *13*(June), 1–17. https://doi.org/10.3389/fonc.2023.1151257

Alam, T. M., Shaukat, K., Khan, W. A., Hameed, I. A., Almuqren, L. A., Raza, M. A., Aslam, M., & Luo, S. (2022). An Efficient Deep Learning-Based Skin Cancer Classifier for an Imbalanced Dataset. *Diagnostics*, *12*(9). https://doi.org/10.3390/diagnostics12092115

Ali, K., Shaikh, Z. A., Khan, A. A., & Laghari, A. A. (2022). Multiclass skin cancer classification using EfficientNets – a first step towards preventing skin cancer. *Neuroscience Informatics*, *2*(4), 100034. https://doi.org/10.1016/j.neuri.2021.100034

Almutairi, A., & Khan, R. U. (2023). Image-Based Classical Features and Machine Learning Analysis of Skin Cancer Instances. *Applied Sciences (Switzerland)*, *13*(13). https://doi.org/10.3390/app13137712

Alotaibi, A., & AlSaeed, D. (2025). Skin Cancer Detection Using Transfer Learning and Deep Attention Mechanisms. *Diagnostics*, *15*(1). https://doi.org/10.3390/diagnostics15010099

Anand, V., Gupta, S., Altameem, A., Nayak, S. R., Poonia, R. C., Khader, A., & Saudagar, J. (2022). *An Enhanced Transfer Learning Based Classification for Diagnosis of Skin Cancer*.

Angurana, N., Rajan, A. P., & Srivastava, I. (2019). Skin Cancer Detection and Classification. *International Journal of Engineering and Management Research*, *9*(2), 111–114. https://doi.org/10.31033/ijemr.9.2.13

Aruk, I., Pacal, I., & Toprak, A. N. (2025). A novel hybrid ConvNeXt-based approach for enhanced skin lesion classification. *Expert Systems with Applications*, *283*(February), 127721. https://doi.org/10.1016/j.eswa.2025.127721

Behara, K., Bhero, E., & Agee, J. T. (2023). Skin Lesion Synthesis and Classification Using an Improved DCGAN Classifier. *Diagnostics*, *13*(16). https://doi.org/10.3390/diagnostics13162635

El-fattah, I. A., Ali, A. M., El-shafai, W., Taha, T. E., & El-samie, F. E. A. (2023). Deep-learning-based super-resolution and classification framework for skin disease detection applications. *Optical and Quantum Electronics*, *March*. https://doi.org/10.1007/s11082-022-04432-x

Gururaj, H. L., Manju, N., Nagarjun, A., Manjunath Aradhya, V. N., & Flammini, F. (2023). DeepSkin: A Deep Learning Approach for Skin Cancer Classification. *IEEE Access*, *11*(April), 50205–50214. https://doi.org/10.1109/ACCESS.2023.3274848

Hao, S., Zhang, L., Jiang, Y., Wang, J., Ji, Z., Zhao, L., & Ganchev, I. (2023). ConvNeXt-ST-AFF: A Novel Skin Disease Classification Model Based on Fusion of ConvNeXt and Swin Transformer. *IEEE Access*, *11*(September), 117460–117473. https://doi.org/10.1109/ACCESS.2023.3324042

Hosny, K. M., Said, W., Elmezain, M., & Kassem, M. A. (2024). Explainable deep inherent learning for multi-classes skin lesion classification. *Applied Soft Computing*, *159*(November 2023), 111624. https://doi.org/10.1016/j.asoc.2024.111624

Hu, Z., Mei, W., Chen, H., & Hou, W. (2024). Multi-scale feature fusion and class weight loss for skin lesion classification. *Computers in Biology and Medicine*, *176*(May), 108594. https://doi.org/10.1016/j.compbiomed.2024.108594

Jain, S., Singhania, U., Tripathy, B., Nasr, E. A., & Aboudaif, M. K. (2021). Deep Learning-Based Transfer Learning for Classification of Skin Cancer. *Sensors*, *21*(8142), 16. https://doi.org/https://doi.org/10.3390/s21238142

Ji, Z., Wang, X., Liu, C., Wang, Z., Yuan, N., & Ganchev, I. (2024). EFAM-Net: A Multi-Class Skin Lesion Classification Model Utilizing Enhanced Feature Fusion and Attention Mechanisms. *IEEE Access*, *September*, 143029–143041. https://doi.org/10.1109/ACCESS.2024.3468612

Kalaiarasan, R., Madhan Kumar, K., Sridhar, S., & Yuvarai, M. (2022). Deep Learning-based Transfer Learning for Classification of Skin Cancer. *Proceedings - International Conference on Applied Artificial Intelligence and Computing, ICAAIC 2022*, 450–454. https://doi.org/10.1109/ICAAIC53929.2022.9792651

Karthik, R., Vaichole, T. S., Kulkarni, S. K., Yadav, O., & Khan, F. (2022). Eff2Net: An efficient channel attention-based convolutional neural network for skin disease classification. *Biomedical Signal Processing and Control*, *73*(October 2021), 103406. https://doi.org/10.1016/j.bspc.2021.103406

Kavitha, C., Priyanka, S., Kumar, M. P., & Kusuma, V. (2024). Skin Cancer Detection and Classification using Deep Learning Techniques. *Procedia Computer Science*, *235*(2023), 2793–2802. https://doi.org/10.1016/j.procs.2024.04.264

Li, Z., Koban, K. C., Schenck, T. L., Giunta, R. E., Li, Q., & Sun, Y. (2022). Artificial Intelligence in Dermatology Image Analysis: Current Developments and Future Trends. *Journal of Clinical Medicine*, *11*(22). https://doi.org/10.3390/jcm11226826

Lilhore, U. K., Sharma, Y. K., Simaiya, S., Alroobaea, R., Baqasah, A. M., Alsafyani, M., & Alhazmi, A. (2025). SkinEHDLF a hybrid deep learning approach for accurate skin cancer classification in complex systems. *Scientific Reports*, *15*(1), 1–32. https://doi.org/10.1038/s41598-025-98205-7

Liu, Wang, X., Liu, H., Zang, X., Li, L., Ji, Z., & Ganchev, I. (2025). FUSCANet: Enhancing Skin Disease Classification through Feature Fusion and Spatial-Channel Attention Mechanisms. *IEEE Access*, *13*(June), 100683–100698. https://doi.org/10.1109/ACCESS.2025.3577740

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., &

Xie, S. (2022). A ConvNet for the 2020s. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11966–11976. https://doi.org/10.1109/CVPR52688.2022.01167

Malik, S. G., Jamil, S. S., Aziz, A., Ullah, S., Ullah, I., & Abohashrh, M. (2024). High-Precision Skin Disease Diagnosis through Deep Learning on Dermoscopic Images. *Bioengineering*, *11*(9), 867. https://doi.org/10.3390/bioengineering11090867

Niino, M., & Matsuda, T. (2021). Age-specific skin cancer incidence rate in the world. *Japanese Journal of Clinical Oncology*, *51*(5), 848–849. https://doi.org/10.1093/jjco/hyab057

Shahzad, K., Wasim, M., Pires, I. M., & Garcia, N. M. (2024). Multi- classification of skin lesions using a deep learning-based convolutional neural network. *Procedia Computer Science*, *241*(2019), 588–593. https://doi.org/10.1016/j.procs.2024.08.085

Su, Q., Hamed, H. N. A., Isa, M. A., Hao, X., & Dai, X. (2024). A GAN-Based Data Augmentation Method for Imbalanced Multi-Class Skin Lesion Classification. *IEEE Access*, *12*(January), 16498–16513. https://doi.org/10.1109/ACCESS.2024.3360215

Tschandl, P. (2021). Artificial intelligence for melanoma diagnosis. *Italian Journal of Dermatology and Venereology*, *156*(3), 289–299. https://doi.org/10.23736/S2784-8671.20.06753-X

Tschandl, P., Rosendahl, C., & Kittler, H. (2018). Data descriptor: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, *5*, 1–10. https://doi.org/10.1038/sdata.2018.161

Wei, M., Wu, Q., Ji, H., Wang, J., Lyu, T., Liu, J., & Zhao, L. (2023). A Skin Disease Classification Model Based on DenseNet and ConvNeXt Fusion. *Electronics (Switzerland)*, *12*(2), 1–19. https://doi.org/10.3390/electronics12020438

Xin, C., Liu, Z., Zhao, K., Miao, L., Ma, Y., Zhu, X., Zhou, Q., Wang, S., Li, L., Yang, F., Xu, S., & Chen, H. (2022). An improved transformer network for skin cancer classification. *Computers in Biology and Medicine*, *149*(June), 105939. https://doi.org/10.1016/j.compbiomed.2022.105939