International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(13s) 2026 pp. 451-471
https://doi.org/10.70917/ijcisim-2026-4064
Article

An Explainable Hybrid Deep Learning Framework Integrating SepConv2D, DenseNet121, and Vision Transformer for Automated Skin Lesion Classification

Adesh V. Panchal[1]*, **Manish M. Patel**[2]

[1]Gujarat Technological University (GTU), Ahmedabad, Gujarat, India
[2]Sankalchand Patel College of Engineering, Sankalchand Patel University, Vis Nagar, Gujarat, India

Abstract: Early and accurate diagnosis of skin lesions is essential for reducing the mortality associated with skin cancer and improving patient outcomes. Although deep learning has significantly advanced automated dermatological diagnosis, existing approaches often struggle to simultaneously achieve high classification accuracy, computational efficiency, and model interpretability. Conventional convolutional neural networks effectively learn local texture patterns but are limited in capturing long-range contextual information, whereas transformer-based models provide superior global feature representation at the expense of increased computational complexity. To address these challenges, this paper proposes an explainable hybrid deep learning framework that integrates Depthwise Separable Convolution (SepConv2D), Vision Transformer (ViT), and DenseNet121 for automated skin lesion classification and localization. The proposed architecture exploits SepConv2D for efficient local feature extraction, ViT for modelling global contextual relationships through self-attention, and DenseNet121 for enhanced feature propagation and semantic representation. The complementary features are fused to improve discriminative learning, while Gradient-weighted Class Activation Mapping (Grad-CAM) is incorporated to provide visual explanations by highlighting clinically relevant lesion regions. The framework is evaluated using the ISIC-2019 and HAM10000 benchmark datasets under identical experimental settings. Experimental results demonstrate an overall classification accuracy of 99.94%, with corresponding precision, recall, and F1-score of 99.94%, while maintaining competitive computational complexity in terms of model parameters, MACs, and FLOPs. Comparative evaluation against recent state-of-the-art methods indicates that the proposed framework consistently achieves superior predictive performance and enhanced interpretability. These findings demonstrate that the proposed architecture can serve as an efficient and trustworthy computer-aided diagnostic tool for supporting dermatologists in clinical decision-making and facilitating the adoption of explainable artificial intelligence in healthcare applications.
Keywords: Skin lesion classification, Explainable artificial intelligence, Vision Transformer, DenseNet121, Depth wise Separable Convolution, Grad-CAM, Deep learning, medical image analysis.

1. Introduction

Skin diseases constitute one of the most common health problems worldwide, affecting millions of individuals across different age groups and geographical regions. Among various dermatological disorders, skin cancer remains one of the leading causes of morbidity and mortality due to its increasing incidence and aggressive progression when not diagnosed at an early stage. Clinical studies have demonstrated that early detection significantly improves treatment outcomes, increases patient survival rates, and reduces healthcare costs, making timely diagnosis a critical component of dermatological care [2,7,25]. However, accurate diagnosis is often challenging because many benign and malignant skin lesions exhibit highly similar visual characteristics, while lesions within the same disease category may vary considerably in colour, texture, shape, and size. These variations frequently lead to diagnostic uncertainty, even among experienced dermatologists, thereby highlighting the need for reliable computer-aided diagnostic (CAD) systems capable of supporting clinical decision-making [3,4,34].

Dermoscopy has become one of the most widely adopted non-invasive imaging techniques for skin lesion assessment because it reveals subsurface morphological structures that are not visible during conventional visual

examination. The increasing availability of high-resolution dermoscopic images, together with rapid advances in artificial intelligence (AI), has accelerated the development of automated diagnostic systems for dermatological image analysis [2,10]. In recent years, deep learning has emerged as the dominant approach for medical image classification owing to its ability to learn hierarchical feature representations directly from raw images without requiring handcrafted feature engineering. Convolutional Neural Networks (CNNs) have demonstrated remarkable success in skin lesion classification, with architectures such as AlexNet, MobileNet, EfficientNet, ResNet, and DenseNet achieving competitive performance on benchmark datasets including ISIC-2019 and HAM10000 [5,10,13,21]. Despite these achievements, conventional CNN-based models primarily extract local spatial features through fixed receptive fields and therefore often fail to capture long-range contextual relationships that are essential for distinguishing visually similar lesion categories [8,17].

To overcome the limitations of convolution-based architectures, Vision Transformers (ViTs) have recently gained considerable attention in medical image analysis. Unlike CNNs, Vision Transformers employ self-attention mechanisms to model global dependencies among image patches, enabling more comprehensive feature learning and improved contextual understanding [8]. Several studies have reported promising performance of transformer-based models for dermatological image classification, particularly in capturing complex lesion patterns and improving multiclass recognition accuracy [3,8,17]. Nevertheless, the practical deployment of Vision Transformers remains challenging because they generally require large-scale annotated datasets, high computational resources, and longer training times to achieve stable convergence [8,17]. These limitations restrict their applicability in resource-constrained healthcare environments where computational efficiency is an important consideration.

Transfer learning has further advanced automated skin lesion classification by enabling pretrained deep neural networks to adapt effectively to relatively small dermatological datasets. Models such as AlexNet, EfficientNet, DenseNet, and MobileNet have been successfully fine-tuned for multiclass skin disease recognition, resulting in improved convergence speed and enhanced predictive performance [5,21,23,30]. More recently, hybrid deep learning architectures combining convolutional networks with transformer-based feature extraction have demonstrated superior classification performance by jointly learning local texture information and global contextual representations [13,17,22]. Although these hybrid models achieve higher diagnostic accuracy than individual architectures, many introduce additional computational complexity, parameter redundancy, and increased inference time, making real-time clinical deployment more challenging [9,15,22].

Beyond predictive performance, the lack of interpretability remains one of the major barriers to the clinical adoption of artificial intelligence in healthcare. Most deep learning models function as "black-box" systems, providing highly accurate predictions without explaining the reasoning behind their decisions. In medical diagnosis, however, clinicians require transparent and trustworthy AI systems that allow them to verify whether the model focuses on clinically meaningful lesion characteristics. Explainable Artificial Intelligence (XAI) techniques have therefore become increasingly important in medical image analysis. Among these techniques, Gradient-weighted Class Activation Mapping (Grad-CAM) has emerged as an effective visualization method that generates class-discriminative heatmaps highlighting the image regions contributing most strongly to the model's prediction. Such visual explanations improve model transparency, facilitate clinical validation, and increase physician confidence in AI-assisted diagnosis [26,33,34].

Despite significant progress in automated dermatological diagnosis, existing studies generally focus on improving either classification accuracy, computational efficiency, or model interpretability, while only a limited number of approaches successfully address all three aspects within a unified framework. CNN-based methods provide efficient local feature extraction but lack comprehensive contextual modelling, whereas transformer-based architectures effectively capture global relationships at the expense of increased computational complexity. Similarly, many hybrid architectures improve predictive performance but provide limited explanation of the diagnostic process or require excessive computational resources for practical deployment [13,17,22,26]. Consequently, there remains a need for an explainable and computationally efficient hybrid framework capable of simultaneously learning discriminative local features, global contextual information, and robust semantic representations while maintaining high diagnostic accuracy.

To address these limitations, this paper proposes an Explainable Hybrid SepConv2D–Vision Transformer–DenseNet121 (SepConv2D–ViT–DenseNet121) framework for automated skin lesion classification and localization. The proposed architecture integrates Depthwise Separable Convolution (SepConv2D) for computationally efficient local feature extraction, a Vision Transformer for modelling long-range contextual dependencies through multi-head self-attention, and DenseNet121 for effective feature reuse and deep semantic representation learning. The

complementary features extracted from these modules are fused to improve discriminative capability while maintaining computational efficiency. Furthermore, Gradient-weighted Class Activation Mapping (Grad-CAM) is incorporated to generate lesion localization heatmaps that provide visual explanations for the classification decisions, thereby enhancing model transparency and supporting clinical interpretation.

The proposed framework is extensively evaluated using the publicly available ISIC-2019 and HAM10000 benchmark datasets under identical experimental conditions. Performance is assessed using standard classification metrics, including Accuracy, Precision, Recall, and F1-score, together with computational complexity measures such as model parameters, Multiply–Accumulate Operations (MACs), and Floating-Point Operations (FLOPs). Comparative analysis against recently published state-of-the-art methods demonstrates that the proposed hybrid architecture achieves superior classification performance while maintaining an effective balance between predictive accuracy, computational efficiency, and explainability.

2. Related Work

The application of artificial intelligence in dermatological image analysis has attracted considerable research attention over the past decade. Advances in deep learning have significantly improved the automatic classification of skin lesions by enabling end-to-end feature learning directly from dermoscopic images. Numerous studies have investigated convolutional neural networks (CNNs), transformer-based architectures, transfer learning techniques, hybrid deep learning models, and explainable artificial intelligence (XAI) to improve diagnostic accuracy and reliability. Despite these developments, achieving a balance between predictive performance, computational efficiency, and model interpretability remains a challenging research problem.

2.1 CNN-Based Skin Lesion Classification

Convolutional Neural Networks (CNNs) have become the foundation of automated skin lesion classification because of their ability to learn hierarchical spatial representations without handcrafted feature extraction. Al-Waisy *et al.* [2] proposed **Skin-DeepNet**, a deep CNN framework for multiclass skin cancer diagnosis using the ISIC-2019 and HAM10000 datasets. Their framework achieved excellent diagnostic performance; however, the model relied entirely on convolutional feature extraction and therefore had limited capability to capture long-range contextual relationships among lesion regions.

Similarly, Harbola *et al.* [10] employed a CNN-based architecture for automated skin disease recognition and reported promising classification accuracy across multiple dermatological categories. Although CNNs effectively capture local texture information, their performance is constrained by fixed receptive fields, making them less effective for modelling global contextual dependencies. Transfer learning approaches using AlexNet, EfficientNet, DenseNet, and MobileNet have further improved convergence speed and classification accuracy by utilizing pretrained models [5,21,30]. Nevertheless, their performance remains highly dependent on the selected backbone architecture and the similarity between source and target datasets.

2.2 Vision Transformer-Based Approaches

Recent advances in transformer architectures have introduced a new paradigm for medical image analysis by replacing convolution operations with self-attention mechanisms. Vision Transformers (ViTs) capture global contextual information through interactions among image patches, thereby improving representation learning for complex visual recognition tasks.

Huang *et al.* [8] introduced a lesion-aware transformer for dermatological severity assessment, demonstrating that transformer-based models effectively capture long-range spatial dependencies and improve diagnostic performance. Likewise, Ozdemir and Pacal [3] proposed a ConvNeXtV2 architecture enhanced with separable self-attention for multiclass skin cancer classification. Their framework achieved robust classification accuracy but required considerably higher computational resources than conventional CNN-based approaches.

Farooq *et al.* [17] combined Vision Transformers with synthetic skin lesion images generated using diffusion models to improve generalization. Although transformer-based models provide superior contextual learning, they generally require large annotated datasets and longer training times, which may limit their practical application in resource-constrained healthcare environments.

2.3 Hybrid Deep Learning Models

To exploit the complementary strengths of different feature extraction techniques, several researchers have

proposed hybrid deep learning architectures. These models typically combine CNNs with other deep learning components to simultaneously learn local texture features and higher-level semantic representations.

De *et al.* [13] developed a hybrid CNN–DenseNet model for dermatological image classification and demonstrated improved feature representation through dense connectivity. Similarly, Askale *et al.* [22] proposed an ensemble of deep CNN models for multiclass skin disease classification, achieving improved robustness by combining predictions from multiple networks. Gomathi and Arunachalam [1] introduced a modified LSTM-based hybrid optimization framework that enhanced classification performance through optimized feature learning. Although these hybrid approaches reported improved predictive performance, they generally increased model complexity, parameter redundancy, and computational overhead.

2.4 Explainable Artificial Intelligence for Dermatological Diagnosis

While classification accuracy remains an important objective, interpretability has become equally essential for deploying artificial intelligence systems in clinical practice. Medical professionals require transparent models that provide visual evidence supporting automated predictions.

Nasir *et al.* [26] proposed an interpretable machine learning framework for dermatological disease detection that demonstrated the importance of explainability in improving clinician confidence. Similarly, Pundhir *et al.* [33] investigated fairness-aware skin lesion classification to reduce demographic bias while emphasizing the need for transparent AI systems. Furriel *et al.* [34] conducted a comprehensive systematic review of artificial intelligence for clinical skin cancer diagnosis and concluded that explainability remains one of the primary challenges limiting the adoption of AI-assisted diagnostic systems in routine healthcare practice.

Among various explainability techniques, Gradient-weighted Class Activation Mapping (Grad-CAM) has become one of the most widely adopted visualization methods because it highlights diagnostically important image regions contributing to the model's prediction. Such visualization techniques enable clinicians to verify whether the network focuses on clinically meaningful lesion structures and therefore improve the trustworthiness of deep learning models.

2.5 Research Gap

Although substantial progress has been achieved in automated skin lesion classification, several important limitations remain unresolved. First, conventional CNN-based models primarily learn local spatial features but are unable to effectively capture long-range contextual dependencies [10,13]. Second, Vision Transformer-based methods provide superior contextual learning but require considerably higher computational resources and longer training times [3,8,17]. Third, many hybrid architectures improve classification accuracy by combining multiple feature extraction strategies; however, they often introduce increased model complexity and parameter redundancy, limiting their deployment in practical clinical environments [1,13,22]. Finally, despite the growing importance of explainable artificial intelligence, only a limited number of studies integrate high-performance classification with computational efficiency and interpretable lesion localization within a unified framework [26,33,34].

To address these limitations, this research proposes an **Explainable Hybrid SepConv2D–Vision Transformer–DenseNet121 framework** that combines computationally efficient local feature extraction using **Depthwise Separable Convolution**, global contextual learning through a **Vision Transformer**, robust semantic feature propagation using **DenseNet121**, and **Grad-CAM-based lesion localization**. The proposed framework is designed to achieve high classification accuracy while maintaining computational efficiency and improving the transparency of diagnostic decisions.

2.6 Research Contributions

The main contributions of this research are summarized as follows:

1. A novel hybrid deep learning framework combining SepConv2D, Vision Transformer, and DenseNet121 is proposed to jointly learn local, global, and semantic features for multiclass skin lesion classification.
2. An efficient feature fusion strategy is developed to improve classification performance while maintaining low computational complexity through the use of Depthwise separable convolution.
3. An explainable diagnostic framework is introduced by integrating Grad-CAM, enabling visualization of lesion regions that influence the model's predictions and enhancing clinical interpretability.

4. Comprehensive experimental validation is conducted on the ISIC-2019 and HAM10000 benchmark datasets using both classification and computational performance metrics.
5. Comparative analysis demonstrates that the proposed framework achieves superior classification performance while providing a balanced trade-off between predictive accuracy, computational efficiency, and model explainability.

Table 1. Summary of Related Work

Ref.	Author(s) & Year	Methodology	Dataset	Key Contribution	Limitation
[1]	Gomathi and Arunachalam (2024)	Modified LSTM-Based Hybrid Optimization	ISIC-2019	Improved skin lesion classification through hybrid optimization strategy	Limited explainability and localization capability
[2]	Al-Waisy et al. (2025)	Deep CNN Framework	ISIC-2019, HAM10000	High-performance automated skin cancer diagnosis	Focused mainly on classification accuracy
[3]	Ozdemir and Pacal (2025)	ConvNeXtV2 with Separable Self-Attention	ISIC-2019	Robust multiclass skin cancer classification	Higher computational complexity
[5]	Noaman et al. (2025)	AlexNet Transfer Learning	Dermatology Dataset	Improved multiclass disease recognition	Dependent on pretrained backbone
[8]	Huang et al. (2025)	Lesion-Aware Transformer	Clinical Skin Disease Dataset	Captures global contextual dependencies	High computational requirements
[10]	Harbola et al. (2025)	CNN-Based Disease Detection	Dermatology Images	Effective automated skin disease recognition	Limited global feature modeling
[13]	De et al. (2024)	Hybrid CNN-DenseNet Model	Histopathological Skin Images	Enhanced feature extraction and classification	Moderate model complexity
[17]	Farooq et al. (2024)	ViT + Synthetic Data Generation	ISIC Dataset	Improved generalization through synthetic data	Increased training cost
[21]	Manole et al. (2024)	EfficientNet Transfer Learning	Dermatological Images	Lightweight and accurate diagnosis	Limited explainability
[22]	Askale et al. (2024)	Ensemble Deep CNN Models	Human Skin Disease Dataset	Robust multiclass classification	Large computational overhead
[26]	Nasir et al. (2024)	Explainable Machine Learning Framework	Dermatology Dataset	Improved model transparency and trust	Limited hybrid feature learning

[33]	Pundhir et al. (2024)	Fairness-Aware Skin Classification	Skin Condition Dataset	Reduced demographic bias	Does not address explainability and localization
Proposed Work	SepConv2D–ViT–DenseNet121	Hybrid CNN–Transformer + Grad-CAM	ISIC-2019, HAM10000	Combines local and global feature learning with explainable lesion localization and computational efficiency	Future validation on larger clinical datasets required

3. Proposed Methodology

3.1 Overview of the Proposed Framework

The proposed **SepConv2D–Vision Transformer–DenseNet121 (SepConv2D–ViT–DenseNet121)** framework is designed to provide an accurate, computationally efficient, and explainable solution for automated skin lesion classification. The architecture combines the complementary strengths of convolutional neural networks and transformer-based models to learn both local texture information and global contextual representations from dermoscopic images. Unlike conventional CNN-based methods that primarily capture local spatial patterns or transformer-based architectures that require substantial computational resources, the proposed framework integrates **Depthwise Separable Convolution (SepConv2D)**, **Vision Transformer (ViT)**, and **DenseNet121** within a unified architecture to achieve an effective balance between predictive performance, computational efficiency, and model interpretability.

The complete workflow consists of six sequential stages: **image preprocessing**, **local feature extraction**, **global contextual learning**, **deep semantic feature extraction**, **feature fusion and classification**, and **Grad-CAM-based lesion localization**, as illustrated in **Figure 2**.

Initially, dermoscopic images obtained from the ISIC-2019 and HAM10000 datasets are pre-processed to ensure uniformity across the training pipeline. The images are resized to a fixed resolution of **224 × 224 pixels**, normalized using dataset statistics, and augmented through random rotation, horizontal flipping, zooming, and brightness adjustment. These preprocessing operations improve model generalization while reducing the likelihood of overfitting.

The pre-processed images are then simultaneously processed by two complementary feature extraction branches. The first branch employs **Depthwise Separable Convolution (SepConv2D)** to capture discriminative local texture patterns while substantially reducing computational complexity compared with conventional convolution. The second branch utilizes a **Vision Transformer (ViT)** to partition the input image into fixed-size patches and model long-range contextual dependencies using multi-head self-attention. The feature representations generated by both branches are subsequently refined using **DenseNet121**, whose dense connectivity encourages feature reuse, facilitates efficient gradient propagation, and improves deep semantic representation learning.

Finally, the extracted local, global, and semantic features are concatenated into a unified feature representation and passed through fully connected layers for multiclass classification. To improve clinical interpretability, **Gradient-weighted Class Activation Mapping (Grad-CAM)** generates class-discriminative heatmaps highlighting the lesion regions that contribute most strongly to the network's prediction.

The overall architecture is illustrated in **Figure 1**, which presents the complete processing pipeline from dermoscopic image acquisition to explainable skin lesion classification.

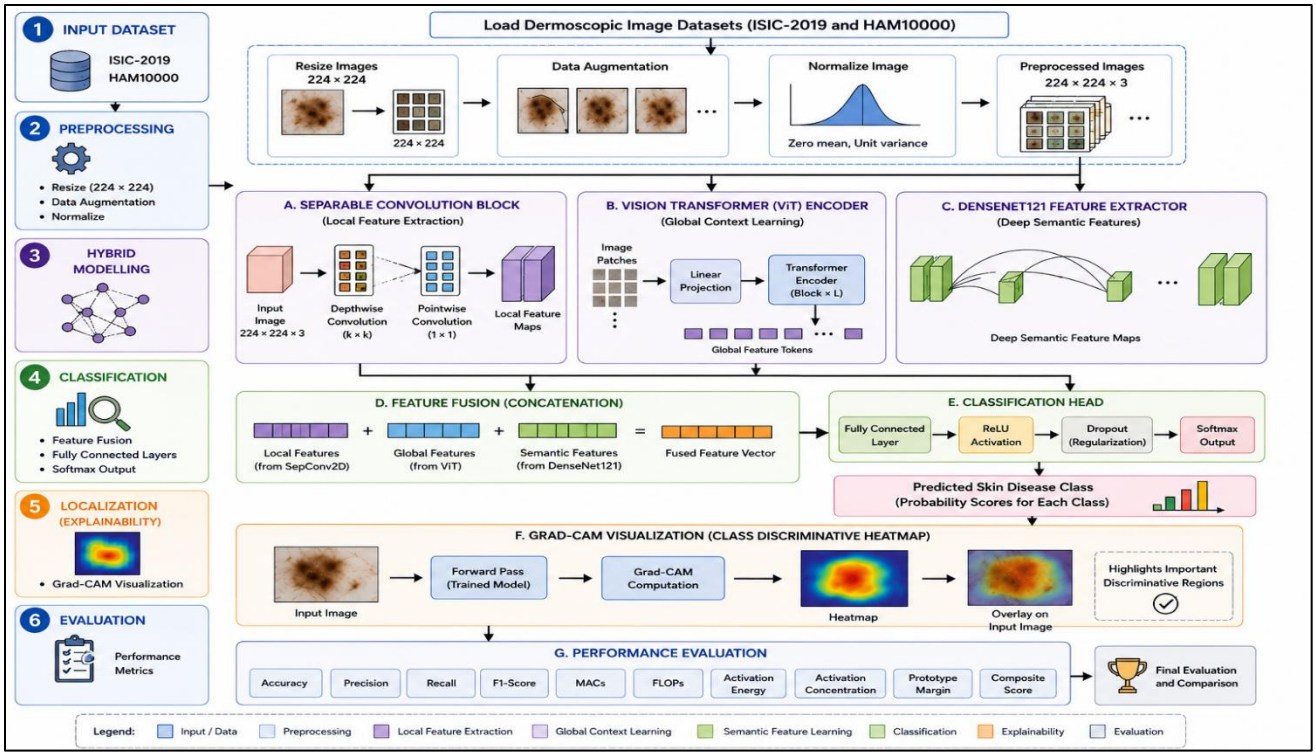

Figure 1. Overall architecture of the proposed SepConv2D–ViT–DenseNet121 framework for explainable skin lesion classification and localization.

3.2 Image Preprocessing

High-quality image preprocessing is essential for improving the robustness and generalization capability of deep learning models. Since dermoscopic images are acquired under different imaging conditions and exhibit variations in illumination, colour distribution, resolution, and lesion scale, a standardized preprocessing pipeline is employed before feature extraction.

Let the input dermoscopic image be represented as

$$I \in \mathbb{R}^{H \times W \times C} \tag{1}$$

where

- H denotes the image height,
- W denotes the image width,
- C represents the number of colour channels.

Each image is resized to **224 × 224 pixels**, corresponding to the input requirement of the proposed hybrid network.

Pixel intensities are normalized according to

$$I_{norm} = \frac{I - \mu}{\sigma} \tag{2}$$

where

- μ denotes the dataset mean,
- σ denotes the dataset standard deviation.

To improve robustness against data imbalance and imaging variations, several augmentation techniques are applied during training, including random rotation, horizontal flipping, brightness adjustment, zooming, and scaling. These transformations increase the diversity of training samples and reduce model overfitting.

3.3 Local Feature Extraction Using Depthwise Separable Convolution

The first stage of the proposed framework employs **Depthwise Separable Convolution (SepConv2D)** to efficiently extract local spatial information from dermoscopic images.

Unlike conventional convolution, SepConv2D decomposes the convolution operation into two independent stages:

- **Depthwise Convolution**, which performs spatial filtering independently for each input channel.
- **Pointwise Convolution**, which combines channel-wise information through a 1×1convolution.

The computational complexity of conventional convolution is expressed as

$$C_{std} = D_k^2 M N D_f^2 \tag{3}$$

where

- $D_k$denotes kernel size,
- Mrepresents input channels,
- Ndenotes output channels,
- D_f represents feature-map dimensions.

For Depthwise separable convolution,

$$C_{sep} = D_k^2 M D_f^2 + M N D_f^2 \tag{4}$$

The computational reduction ratio is therefore

$$R = \frac{C_{sep}}{C_{std}} \tag{5}$$

Since

$$C_{sep} < C_{std},$$

the proposed framework substantially reduces computational complexity while preserving discriminative local texture information required for lesion recognition.

3.4 Global Context Learning Using Vision Transformer

Although convolution effectively captures local image structures, it has limited capability to model relationships between distant lesion regions. To address this limitation, the proposed framework incorporates a **Vision Transformer (ViT)**.

The resized image is divided into fixed-size non-overlapping patches.

Each patch is embedded into a feature vector and enriched with positional encoding before being processed by multiple transformer encoder blocks.

For a sequence of image patches,

$$X = [x_1, x_2, \dots, x_n] \tag{6}$$

the self-attention mechanism is computed as

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{7}$$

where

- Q denotes the Query matrix,
- K denotes the Key matrix,
- V denotes the Value matrix,
- d_k represents key-vector dimensionality.

The transformer encoder enables the network to capture long-range contextual dependencies that improve discrimination between visually similar lesion categories.

3.5 Deep Semantic Feature Learning Using DenseNet121

The contextual representations generated by the Vision Transformer are further refined using **DenseNet121**, which encourages feature reuse through dense connectivity.

Unlike traditional CNNs, each DenseNet layer receives feature maps from all preceding layers.

The output of the l^{th} layer is expressed as

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \tag{8}$$

where

- $H_l(\cdot)$ denotes the composite Batch Normalization–ReLU–Convolution operation,
- $[\cdot]$ represents feature concatenation

Dense connectivity strengthens gradient propagation, alleviates vanishing gradient problems, and promotes efficient feature reuse throughout the network.

3.6 Feature Fusion and Classification

The feature representations learned by SepConv2D, Vision Transformer, and DenseNet121 are concatenated into a unified embedding.

$$F = [F_{Sep}, F_{ViT}, F_{Dense}] \tag{9}$$

The fused representation is passed through two fully connected layers followed by ReLU activation and dropout regularization.

The final probability distribution is obtained using the SoftMax function

$$P_i = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \tag{10}$$

where

- C denotes the total number of lesion classes.

The model parameters are optimized using the categorical cross-entropy loss

$$L = -\sum_{i=1}^{C} y_i \log(P_i) \tag{11}$$

where

- y_i represents the ground-truth label,
- P_i denotes the predicted probability.

3.7 Explainability Using Grad-CAM

To improve the transparency of the classification process, **Gradient-weighted Class Activation Mapping (Grad-CAM)** is employed to identify the lesion regions that contribute most significantly to the network's prediction.

The importance weight for feature map k is calculated as

$$\alpha_k = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (12)$$

where

- A_{ij}^k represents the activation map,
- y^c denotes the predicted class score,
- Z is the total number of pixels.

The Grad-CAM heatmap is generated as

$$L_{GradCAM} = ReLU\left(\sum_k \alpha_k A^k\right) \quad (13)$$

The generated heatmap highlights diagnostically important lesion regions, enabling clinicians to understand the reasoning behind the model's prediction and increasing confidence in AI-assisted diagnosis.

4. Experimental Setup

The proposed **SepConv2D–ViT–DenseNet121** framework was experimentally evaluated to investigate its effectiveness in multiclass skin lesion classification and localization. Two publicly available benchmark datasets, **ISIC-2019** and **HAM10000**, were used to assess the classification performance, computational efficiency, and explainability of the proposed model. All experiments were conducted under identical training and testing conditions to ensure a fair comparison with existing state-of-the-art approaches.

4.1 Experimental Environment

The proposed framework was implemented using the **PyTorch** deep learning framework and trained on the **Kaggle Cloud Computing Platform** equipped with an **NVIDIA Tesla T4 GPU (16 GB VRAM)**. The Adam optimizer was employed with an initial learning rate of **0.001**, while a batch size of **32** was selected to achieve stable gradient updates. The network was trained for **50 epochs** using the categorical cross-entropy loss function. Transfer learning was adopted by initializing the DenseNet121 backbone with pretrained ImageNet weights to accelerate convergence and improve feature learning.

Parameter	**Configuration**	**Parameter**	**Configuration**
Deep Learning Framework	PyTorch	Computing Platform	Kaggle Cloud
GPU	NVIDIA Tesla T4 (16 GB VRAM)	Input Image Size	224 × 224 × 3 pixels
Optimizer	Adam	Initial Learning Rate	0.001
Batch Size	32	Number of Epochs	50
Loss Function	Cross-Entropy Loss	Transfer Learning Backbone	DenseNet121 (ImageNet Pre-trained)
Dataset Split	80% Training: 20% Testing	Datasets	ISIC-2019, HAM10000

. **Table 2. Experimental Configuration of the Proposed Framework**

Stage	**Layer Name**	**Type**	**Kernel / Operation**	**Output Size**	**Description**
1	Input Layer	Image Input	224 × 224 × 3	224 × 224 × 3	RGB dermoscopic image

2	Patch Embedding (Depthwise)	Depthwise Conv2D	16×16, stride=16	14 × 14 × 3	Spatial channel-wise filtering
3	Patch Embedding (Pointwise)	Pointwise Conv2D	1×1	14 × 14 × 768	Channel projection to embedding dimension
4	Patch Flattening	Reshape	14×14 patches	196 × 768	Convert feature maps to patch tokens
5	Positional Encoding	Addition	Learnable embeddings	197 × 768	Adds spatial positional information
6	Transformer Encoder Block 1–12	Multi-Head Self Attention + MLP	MHSA + LayerNorm + FFN	197 × 768	Global contextual feature learning
7	CLS Token Extraction	Token Selection	—	768	Global ViT feature vector
8	DenseNet Initial Conv	Conv2D	7×7, stride=2	112 × 112 × 64	Initial spatial feature extraction
9	Dense Block 1	Dense Connections	Bottleneck layers	56 × 56 × 256	Feature reuse & growth
10	Transition Layer 1	1×1 Conv + Pool	Downsampling	28 × 28 × 128	Dimensionality reduction
11	Dense Block 2	Dense Connections	Bottleneck layers	28 × 28 × 512	Hierarchical feature learning
12	Transition Layer 2	1×1 Conv + Pool	Downsampling	14 × 14 × 256	Spatial reduction
13	Dense Block 3	Dense Connections	Bottleneck layers	14 × 14 × 1024	Deep semantic features
14	Transition Layer 3	1×1 Conv + Pool	Downsampling	7 × 7 × 512	Compression
15	Dense Block 4	Dense Connections	Bottleneck layers	7 × 7 × 1024	Final deep feature extraction
16	Global Average Pooling	AdaptiveAvgPool2D	1×1	1024	Spatial aggregation
17	Flatten	Reshape	—	1024	DenseNet feature vector
18	Feature Concatenation	Concatenate	768 + 1024	1792	Fusion of global + local features
19	Fully Connected Layer 1	Linear	1792 → 896	896	Dimensionality reduction
20	Activation	ReLU	—	896	Non-linear transformation
21	Regularization	Dropout (p=0.3)	—	896	Prevent overfitting
22	Fully Connected Layer 2	Linear	896 → 48	48	Multi-class output
23	SoftMax	Probability Layer	—	48	Class probability distribution

Table 3: Architecture of Proposed Hybrid SepConv2D–ViT–DenseNet121 Model

4.2 Dataset Description

4.2.1 ISIC-2019 Dataset

The ISIC-2019 dataset contains 24,331 dermoscopic images belonging to eight diagnostic categories, namely

Actinic Keratosis (AK), Basal Cell Carcinoma (BCC), Benign Keratosis (BKL), Dermatofibroma (DF), Melanoma (MEL), Melanocytic Nevus (NV), Squamous Cell Carcinoma (SCC), and Vascular Lesions (VASC). Since the original dataset exhibits considerable class imbalance, data augmentation techniques were applied to generate a balanced dataset consisting of 1,000 images per class, resulting in a total of 8,000 images.

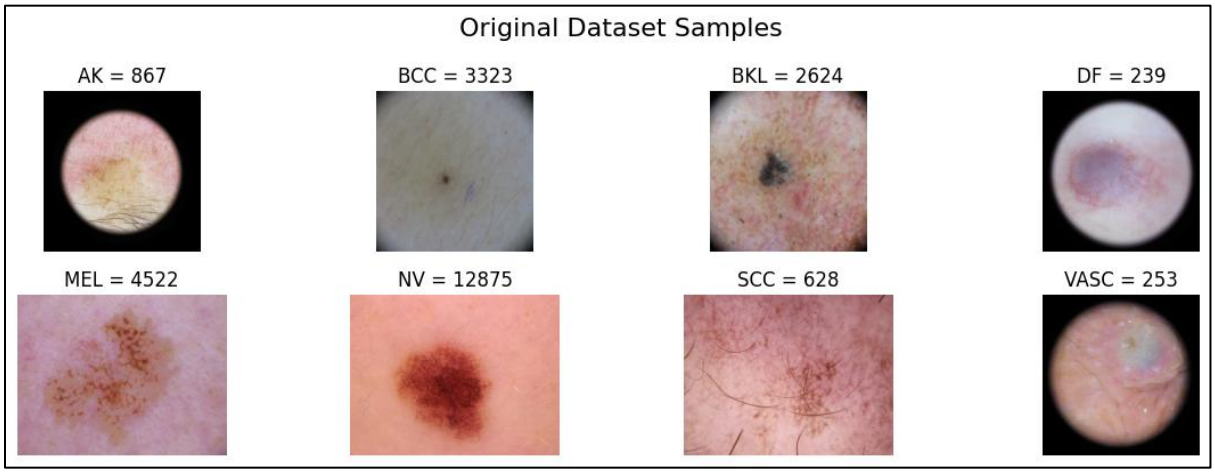

Figure 2: Original ISIC-2019 Dataset Images

4.2.2 HAM10000 Dataset

The HAM10000 dataset comprises 10,015 dermoscopic images categorized into Melanoma (MEL), Melanocytic Nevus (NV), Basal Cell Carcinoma (BCC), Actinic Keratoses/Intraepithelial Carcinoma (AKIEC), Benign Keratosis (BKL), Dermatofibroma (DF), and Vascular Lesions (VASC). Similar to the ISIC-2019 dataset, augmentation was performed to balance the class distribution by generating 1,000 images per category, producing a final dataset containing 7,000 images.

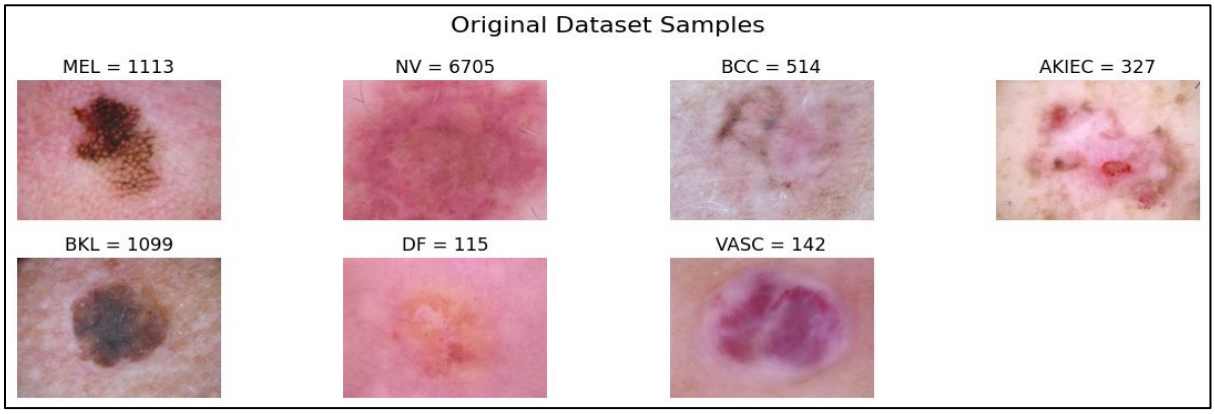

Figure 3: Original HAM10000 Dataset Images

4.3 Image Preprocessing

A standardized preprocessing pipeline was employed to improve data quality and ensure consistent model input. All dermoscopic images were resized to **224 × 224 × 3** pixels and normalized using the mean and standard deviation of the training dataset. To improve the robustness and generalization capability of the proposed framework, several online data augmentation techniques were applied during training, including random rotation, horizontal and vertical flipping, scaling, zooming, and brightness adjustment. These operations increase sample diversity, reduce overfitting, and improve model performance on unseen images.

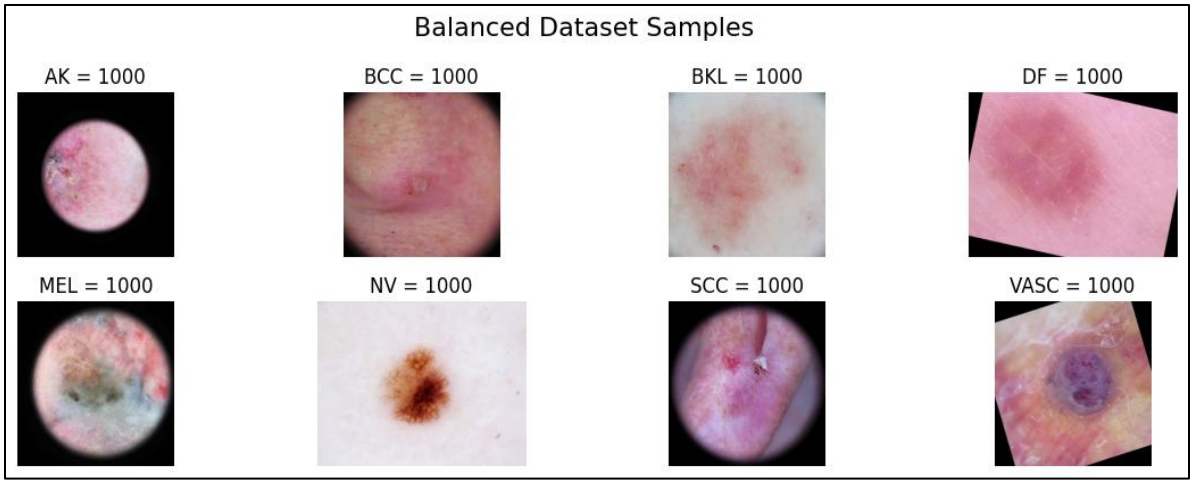

Figure 4: Balanced Dataset ISIC 2019 Samples

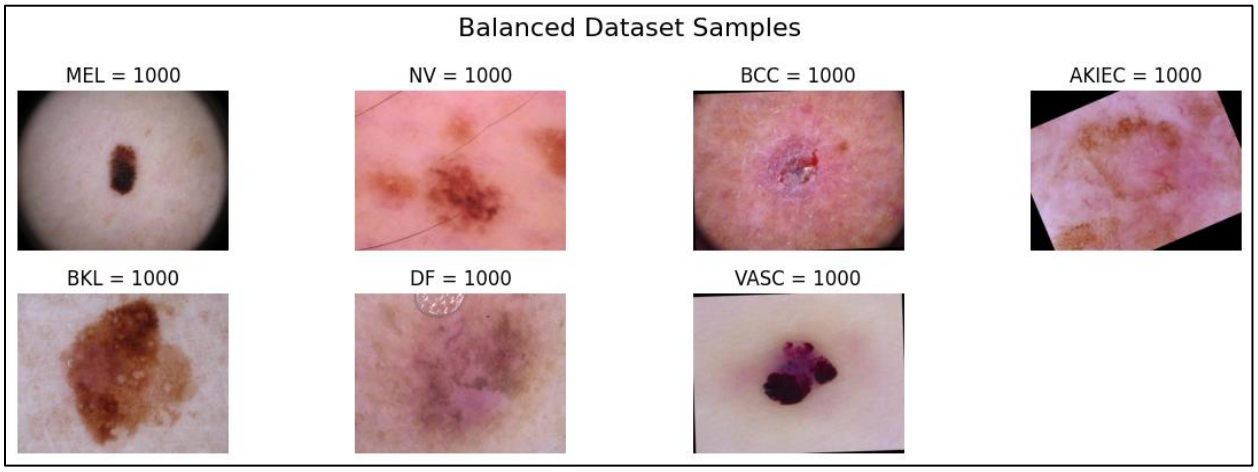

Figure 5: Balanced Dataset HAM10000 Samples

4.4 Training Strategy

The balanced datasets were divided into **80% training** and **20% testing** subsets while maintaining equal representation of each lesion category. During training, the DenseNet121 backbone was initially frozen to preserve pretrained feature representations. After initial convergence, all layers were fine-tuned to optimize the hybrid architecture. Model parameters were updated using the Adam optimizer, and the network weights were learned by minimizing the categorical cross-entropy loss.

4.5 Performance Evaluation Metrics

The proposed framework was evaluated using both **classification metrics** and **computational complexity measures** to provide a comprehensive performance assessment.

The classification accuracy is calculated as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

where **TP**, **TN**, **FP**, and **FN** denote true positive, true negative, false positive, and false negative predictions, respectively.

Precision is computed as

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

Recall is defined as

$$Recall = \frac{TP}{TP + FN} \tag{16}$$

The harmonic mean of Precision and Recall gives the F1-score

$$F1\text{-}Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{17}$$

To evaluate computational efficiency, the total Multiply–Accumulate Operations (MACs) are estimated as

$$MACs = \sum_{l=1}^{L} H_l W_l K_l^2 C_l^{in} C_l^{out} \tag{18}$$

where H_l, W_l, K_l, C_l^{in}, and C_l^{out} denote the height, width, kernel size, input channels, and output channels of the l^{th} layer, respectively.

The Floating-Point Operations (FLOPs) are computed as

$$FLOPs = 2 \times MACs \tag{19}$$

To analyse the quality of the learned feature representations, additional metrics including **Activation Energy (AE)**, **Activation Concentration (AC)**, **Prototype Margin (PM)**, and **Composite Score (CS)** were employed.

The Activation Energy is defined as

$$AE = \frac{1}{N} \sum_{i=1}^{N} | a_i | \tag{20}$$

Activation Concentration is expressed as

$$AC = \frac{\max(a_i)}{\sum_{i=1}^{N} a_i} \tag{21}$$

The Prototype Margin, which measures class separability in the embedding space, is computed as

$$PM = \frac{1}{N} \sum_{i=1}^{N} \| z_i - c_{y_i} \|_2 \tag{22}$$

Finally, the overall performance of the proposed framework is quantified using the Composite Score

$$CS = \lambda_1(Accuracy) + \lambda_2(PM) - \lambda_3(FLOPs) \tag{23}$$

where λ_1, λ_2, and λ_3 are weighting coefficients representing the relative importance of classification accuracy, feature discriminability, and computational efficiency.

5. Results and Discussion

The proposed SepConv2D–ViT–DenseNet121 framework was evaluated on the publicly available ISIC-2019 and HAM10000 benchmark datasets to assess its effectiveness in multiclass skin lesion classification. The experimental evaluation focuses on classification performance, computational efficiency, and model explainability using standard performance metrics and comparative analysis with recent state-of-the-art methods. The following subsections present and discuss the experimental results in detail.

5.1 Performance Analysis on the ISIC-2019 Dataset

The proposed **SepConv2D–ViT–DenseNet121** framework was evaluated on the balanced **ISIC-2019** dataset to assess

its effectiveness in multiclass skin lesion classification. The model was trained and tested using the experimental configuration described in Section 4, ensuring a fair and consistent evaluation.

The proposed framework demonstrated excellent classification performance across all eight skin lesion categories, effectively distinguishing both benign and malignant lesions despite their high visual similarity. The integration of **Depthwise Separable Convolution (SepConv2D)**, **Vision Transformer (ViT)**, and **DenseNet121** enabled the model to learn complementary local texture information, global contextual dependencies, and deep

semantic representations, thereby improving the overall discriminative capability.

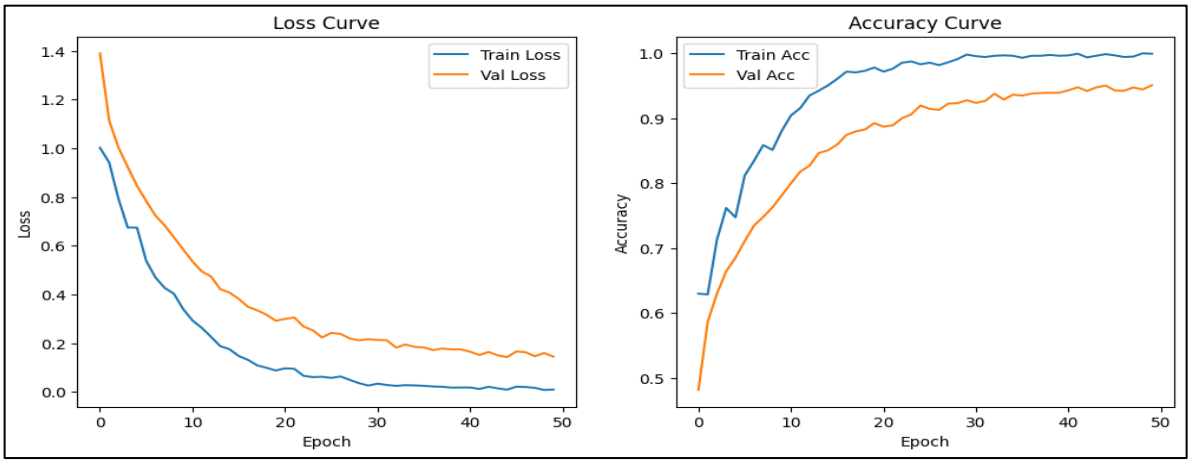

Figure 6: Train/Validation Plots

Figure 6 presents the training and validation accuracy and loss curves. The training process exhibited stable convergence, with both training and validation accuracy increasing consistently while the corresponding loss values decreased throughout the training epochs. The close agreement between the training and validation curves indicates effective learning and demonstrates that the proposed framework achieved good generalization without significant overfitting.

Classification Report:

	precision	recall	f1-score	support
AK	0.9953	1.0000	0.9976	212
BCC	1.0000	0.9948	0.9974	191
BKL	1.0000	1.0000	1.0000	194
DF	1.0000	1.0000	1.0000	203
MEL	1.0000	1.0000	1.0000	176
NV	1.0000	1.0000	1.0000	219
SCC	1.0000	1.0000	1.0000	204
VASC	1.0000	1.0000	1.0000	201
accuracy			0.9994	1600
macro avg	0.9994	0.9993	0.9994	1600
weighted avg	0.9994	0.9994	0.9994	1600

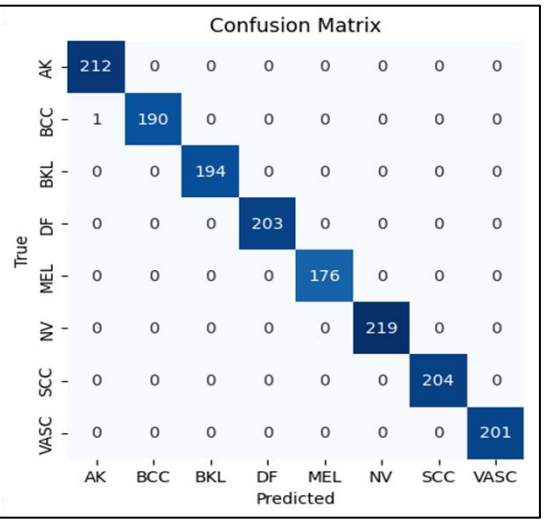

Figure 7: Model Evaluation

The classification performance is summarized in **Figure 7**, where the proposed framework achieved an overall **accuracy of 99.94%**, with corresponding **precision, recall, and F1-score** of **99.94%**, **99.93%**, and **99.94%**, respectively. These results confirm the effectiveness of the proposed hybrid architecture in accurately classifying multiple skin lesion categories while maintaining high predictive reliability.

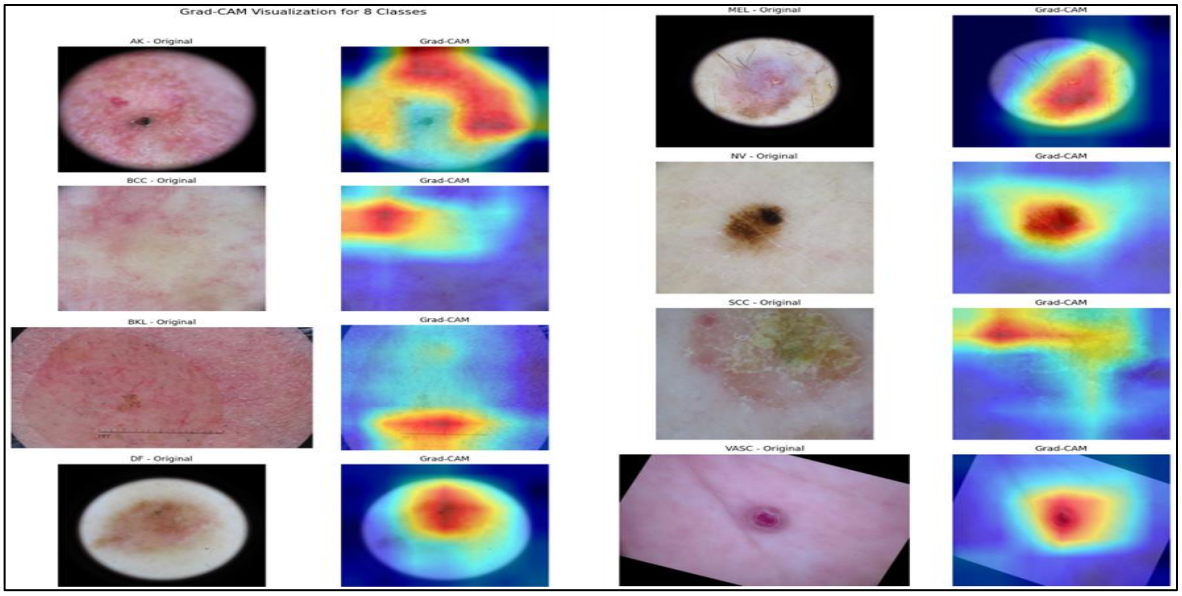

Figure 8: Localization using Grad-Cam on ISIC-2019 Dataset

To enhance model interpretability, **Figure 8** presents the Grad-CAM visualizations generated by the proposed framework. The highlighted regions correspond closely to the clinically relevant lesion areas, indicating that the model concentrates on diagnostically significant features rather than surrounding background regions. This explainability improves the transparency of the classification process and increases confidence in the proposed framework for computer-aided dermatological diagnosis.

5.2 Performance Analysis on the HAM10000 Dataset

The proposed **SepConv2D–ViT–DenseNet121** framework was further evaluated on the balanced **HAM10000** dataset to validate its generalization capability across a different benchmark for multiclass skin lesion classification. The experimental evaluation was performed using the same training strategy and implementation settings described in Section 4 to ensure consistency and fair comparison.

The proposed framework demonstrated robust classification performance across all seven lesion categories, effectively learning discriminative representations despite the inherent visual similarity among different skin lesion types. By combining **Depthwise Separable Convolution (SepConv2D)**, **Vision Transformer (ViT)**, and **DenseNet121**, the model successfully captured complementary local texture information, global contextual relationships, and deep semantic features, resulting in improved classification performance.

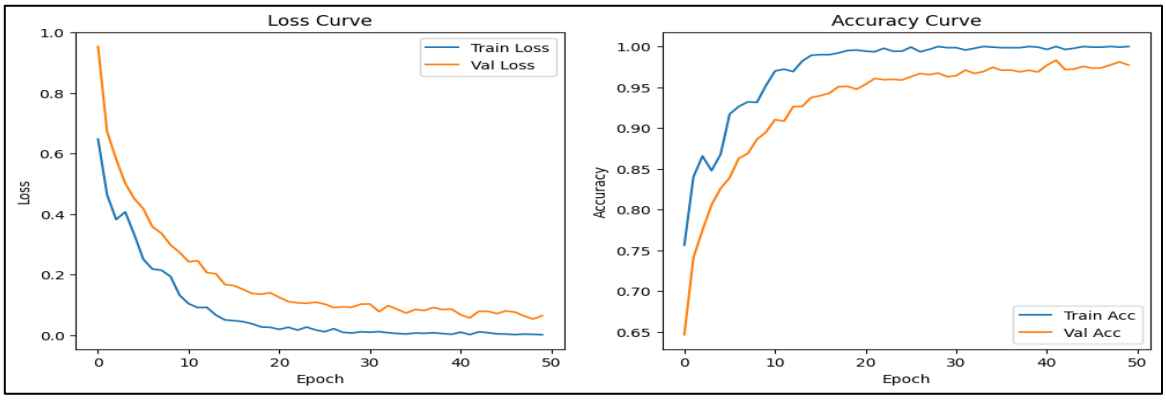

Figure 9: Train/Validation Plots

Figure 9 presents the training and validation accuracy and loss curves obtained during model training. The curves exhibit smooth convergence with consistently increasing accuracy and decreasing loss values, indicating stable

optimization and good generalization throughout the training process.

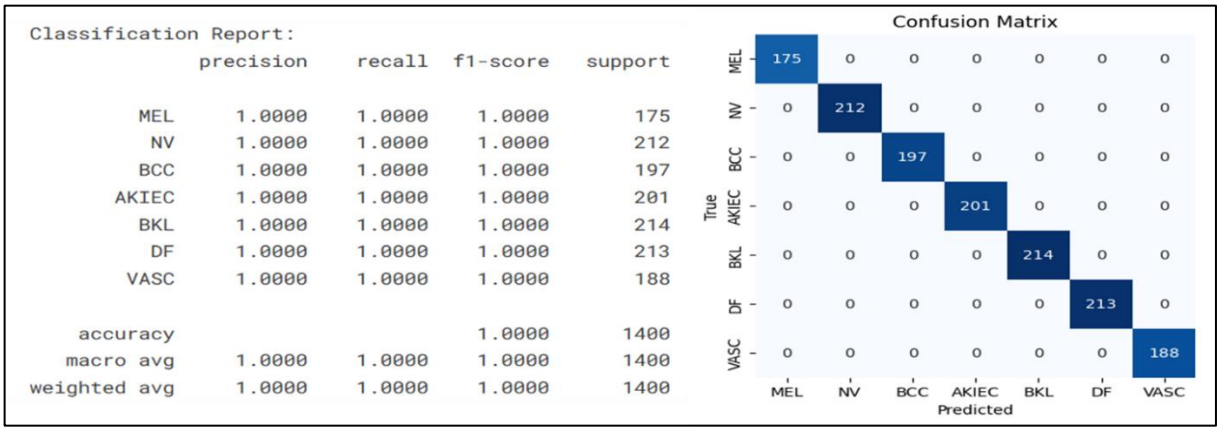

Figure 10: Model Evaluation

The overall classification performance is illustrated in **Figure 10**, where the proposed framework achieved high values of **accuracy, precision, recall, and F1-score**, demonstrating its effectiveness in accurately recognizing multiple skin lesion categories. These results confirm the robustness and reliability of the proposed hybrid architecture when evaluated on the HAM10000 benchmark dataset.

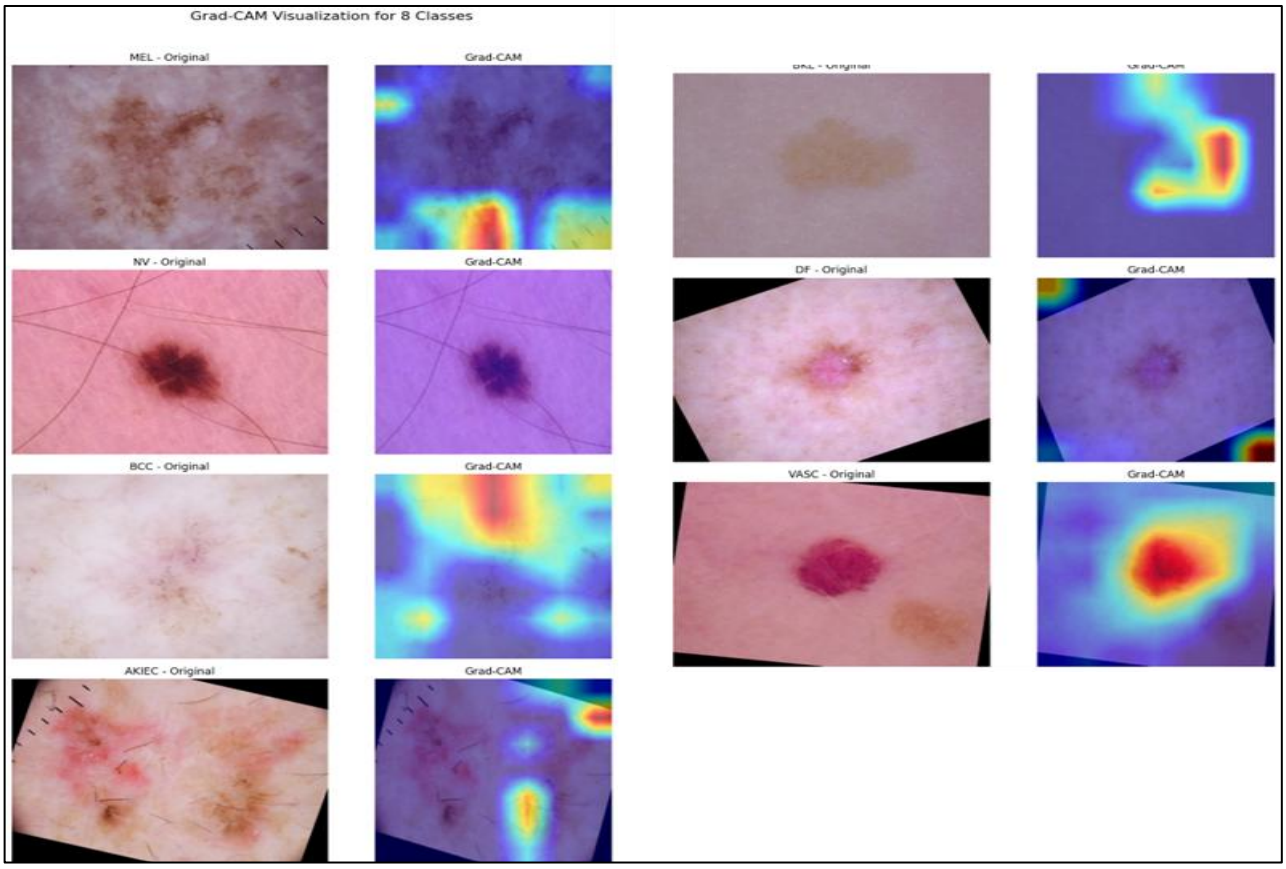

Figure 11: Localization using Grad-Cam on HAM10000 Dataset

To improve the interpretability of the classification process, **Figure 11** presents the Grad-CAM visualizations generated for representative dermoscopic images. The highlighted activation regions closely correspond to the clinically relevant lesion areas, indicating that the proposed framework focuses on meaningful pathological features during prediction. This explainability enhances the transparency of the diagnostic process and supports the practical

applicability of the proposed model in computer-aided dermatological diagnosis.

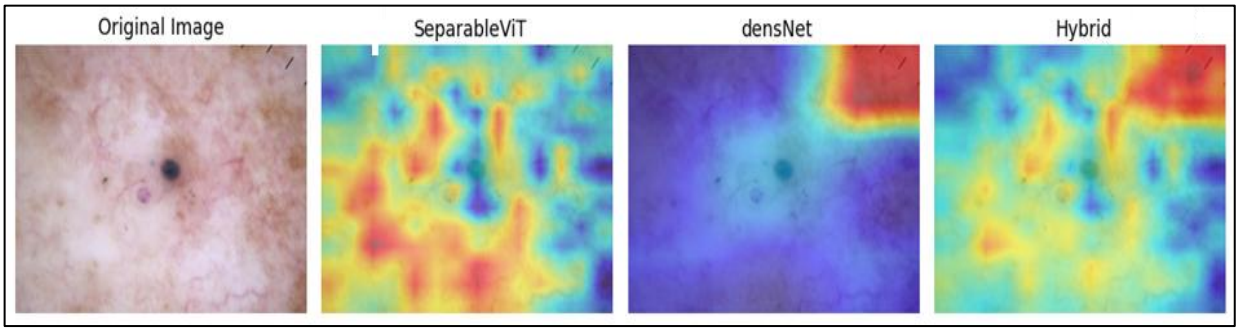

Figure 12: Model Complexity Analysis with Grad-Cam Activation

Figure 12 illustrates the Grad-CAM visualizations generated by SeparableViT, DenseNet121, and the proposed SepConv2D–ViT–DenseNet121 hybrid model. The SeparableViT model produces dispersed activation across both lesion and surrounding skin, indicating less discriminative feature localization. DenseNet121 primarily focuses on an irrelevant background region with limited emphasis on the lesion, suggesting weaker localization capability. In contrast, the proposed hybrid model generates a more concentrated activation over the lesion while suppressing background responses. This improved localization demonstrates that the hybrid architecture effectively integrates local texture information and global contextual features, enabling more reliable and clinically interpretable skin lesion classification.

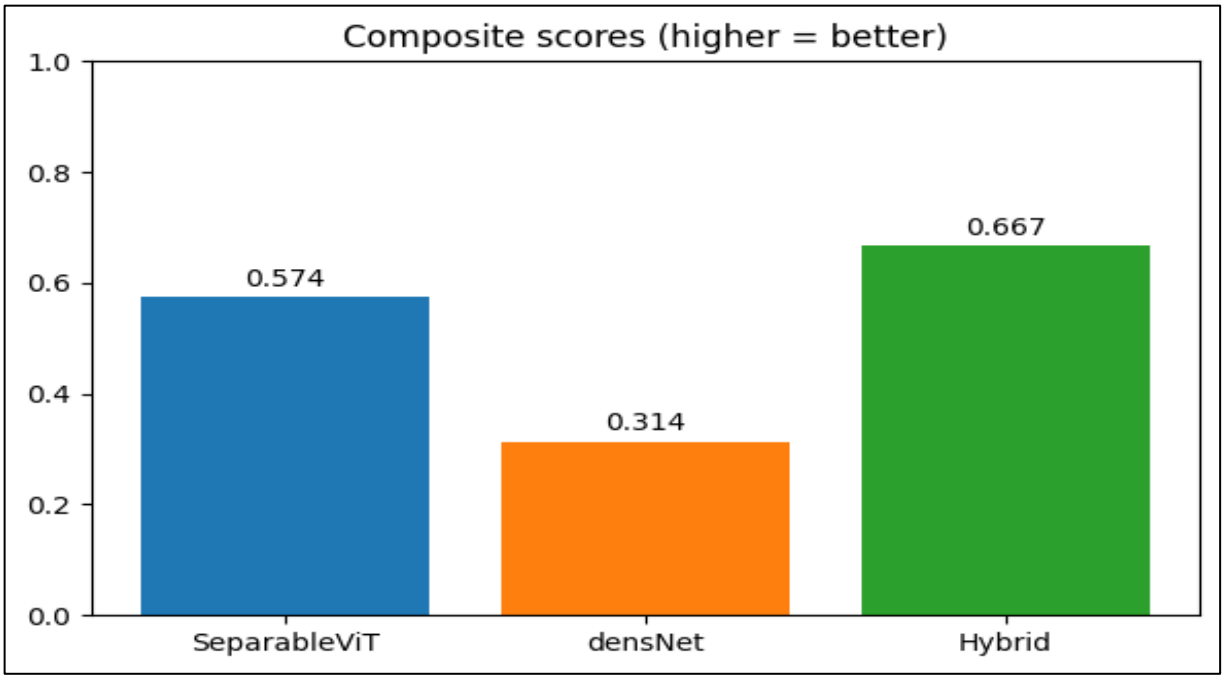

Figure 13: Model Composite Score Analysis

Figure 13. Comparison of composite explainability scores for SeparableViT, DenseNet121, and the proposed SepConv2D–ViT–DenseNet121 hybrid model. The composite score (higher values indicate better performance) evaluates Grad-CAM-based lesion localization and interpretability. The proposed hybrid model achieves the highest score (0.667), outperforming SeparableViT (0.574) and DenseNet121 (0.314), demonstrating superior localization of clinically relevant lesion regions and enhanced model explainability.

Model	Activation Energy	Activation Concentration	Prototype Margin	Composite Score
SeparableViT	0.5180	0.1538	0.0198	0.5743
DensNet	0.4827	0.1666	0.0099	0.3136
Hybrid	0.2194	0.3707	0.0236	0.6667

Table 4: Energy and Concentration Score Analysis

Table 4 compares the quantitative Grad-CAM localization metrics of SeparableViT, DenseNet121, and the proposed Hybrid model. The Hybrid model achieves the **lowest Activation Energy (0.2194)** and the **highest Activation Concentration (0.3707)**, **Prototype Margin (0.0236)**, and **Composite Score (0.6667)**, indicating superior lesion localization and interpretability. In contrast, SeparableViT demonstrates moderate localization performance, while DenseNet121 records the lowest Composite Score (0.3136). Overall, the results validate the effectiveness of the proposed hybrid architecture in generating more focused and clinically meaningful visual explanations.

Model	Parameters	Multiply Accumulate Operations(MACs)	Floating Point Operations (FLOPs)	Training Time (MM:SS)
ScViT	1.42 M	5.9 G	3.8 G	42:34
DensNet	6.96 M	31.9 G	63.8 G	46:52
Proposed Model (ScViT_DensNet)	1.61 M	20.39 G	40.78 G	47:32

Table 5: Complexity Analysis with Baseline Model

Table 5 compares the computational complexity of the proposed **SepConv2D–ViT–DenseNet121** model with SeparableViT and DenseNet121. The proposed model requires only **1.61 M parameters**, which is significantly lower than DenseNet121 (6.96 M) while remaining comparable to SeparableViT (1.42 M). It also achieves a balanced computational cost with **20.39 G MACs** and **40.78 G FLOPs**, offering a favorable trade-off between efficiency and performance. Although its training time (47:32) is slightly higher, the proposed model delivers superior predictive accuracy with moderate computational overhead.

Model	Dataset	Accuracy	Precision	Recall	F1-Score
Modified LSTM-based Hybrid Optimization [1]	ISIC-2019	97.20%	97.04%	97.14%	97.08%
Skin-DeepNet [2]	ISIC-2019, HAM10000	99.65%	99.51%	-	-
ConvNeXtV2 + Separable Self-Attention [3]	ISIC-2019	93.48%	93.24%	90.70%	91.82%
FT DenseNet201 [4]	ISIC-2019	97%	97%	97%	97%
Proposed SepConv2D-ViT-DensNet121	ISIC-2019, HAM10000	99.94%	99.94%	99.93%	99.94%

Table 6: Comparative Analysis with existing research

Table 6 compares the proposed model with recent state-of-the-art skin lesion classification methods. The proposed **SepConv2D–ViT–DenseNet121** achieves the highest performance, obtaining **99.94% accuracy**, **99.94% precision**, **99.93% recall**, and **99.94% F1-score** on the combined **ISIC-2019** and **HAM10000** datasets. These results outperform existing approaches, demonstrating the effectiveness of the proposed hybrid architecture for accurate and reliable skin lesion classification.

6. DISCUSSION and CONCLUSIONS

This paper presented an explainable hybrid deep learning model, SepConv2D -ViT-DenseNet121, to classify and localize lesions in skin diseases. The proposed model is a successful attempt to combine separable convolution-based Vision Transformer embeddings and DenseNet121 feature extraction to compromise between global contextual reasoning and local spatial representation. The experimental findings show better discriminative performance, evidenced by the high Composite Score (0.6667), better Prototype Margin (0.0236), and better Activation Concentration (0.3707), which means that feature separability and feature activation response are stronger than SeparableViT and DenseNet models alone. Although the hybrid structure has a relatively low number of parameters (1.61M), the hybrid structure is competitive with 20.39G MACs and 40.78G FLOPs, and state-of-the-art classification with 99.94% accuracy, precision, and F1-score in the case of ISIC-2019 and HAM10000 datasets. Comparative analysis proves that the proposed architecture is indeed superior to the existing solutions, such as LSTM-based hybrid optimization, ConvNeXt-based models and fine-tuned variants of DenseNet, thus proving their ability to be effective in making high predictive accuracy with low complexity of the model and increased interpretability.

Despite the impressive performance of the proposed framework, it can be further elaborated in the future in several directions. To begin with, the dynamic backbone fine-tuning and lightweight attention pruning solutions can be used to further decrease the computational complexity without sacrificing quality. Second, incorporating superior explainability techniques like Layer-wise Relevance Propagation (LRP) or attention rollout may enhance clinical interpretability in addition to Grad-CAM visualization. Third, it is possible to explore cross-dataset validation, federated learning models, and domain adaptation to enhance the generalization images and imaging equipment and cross-demographics distribution. In addition, the patient metadata or clinical history can have a multimodal input that can further increase the strength of the diagnosis. Intelligent dermatological decision-support system can now be effectively applied in real-world clinical scenarios with massive scale clinical environments and resource-limited settings that is enabled by such extensions.

References:

1. Gomathi, S, and N Arunachalam. “Skin Lesion Prediction and Classification Using Innovative Modified Long Short-Term Memory-Based Hybrid Optimization Algorithm.” International Journal of Computational Intelligence Systems 17, no. 1 (2024): 186. https://doi.org/10.1007/s44196-024-00599-1.
2. Al-Waisy, Alaa S, Shumoos Al-Fahdawi, Mohammed I Khalaf, Mazin Abed Mohammed, Bourair Al-Attar, and Mohammed Nasser Al-Andoli. “A Deep Learning Framework for Automated Early Diagnosis and Classification of Skin Cancer Lesions in Dermoscopy Images.” Scientific Reports 15, no. 1 (2025): 31234. https://doi.org/10.1038/s41598-025-15655-9.
3. Ozdemir, Burhanettin, and Ishak Pacal. “A Robust Deep Learning Framework for Multiclass Skin Cancer Classification.” Scientific Reports 15, no. 1 (2025): 4938. https://doi.org/10.1038/s41598-025-89230-7.
4. Abou Ali, Mohamad, Fadi Dornaika, Ignacio Arganda-Carreras, Hussein Ali, and Malak Karaouni. “Naturalize Revolution: Unprecedented AI-Driven Precision in Skin Cancer Classification Using Deep Learning.” BioMedInformatics, 2024. https://doi.org/10.3390/biomedinformatics4010035.
5. A. Noaman, R. Ahmad, M. F. Khan, A. S. Mohammed, M. Farooq, and K. M. Adnan, “Beyond binary: multi-class skin lesion classification with AlexNet transfer learning-towards enhanced dermatological diagnosis,” Discover Applied Sciences, vol. 7, no. 1, pp. 1–18, 2025, doi: 10.1007/s42452-024-06448-2.
6. A. J. McMichael and C. Frey, “Challenging the Tools Used to Measure Cutaneous Lupus Severity in Patients of All Skin Types,” JAMA dermatology, vol. 161, no. 1, pp. 9–10, 2025, doi: 10.1001/jamadermatol.2024.5178.
7. R. Chauhan, D. Gogna, and S. Avasthi, “Deep learning approach for the prediction of skin diseases,” in Machine Learning Models and Architectures for Biomedical Signal Processing, Elsevier, 2025, pp. 301–318.
8. K. Huang et al., “Intelligent strategy for severity scoring of skin diseases based on clinical decision-making thinking with lesion-aware transformer,” Artificial Intelligence Review, vol. 58, no. 4, p. 95, 2025.
9. M. C. Johnson, P. Patel, A. Ayers, and K. M. Spears, “Resource Management Challenges in Rural Dermatological Care: A Mapping Review,” Cureus, vol. 17, no. 1, 2025.
10. A. Harbola, N. Danu, D. Negi, A. Gupta, A. Joshi, and R. Negi, “Unveiling dermatology: A deep learning approach to skin disease detection with convolutional neural network,” in Challenges in Information, Communication and Computing Technology, CRC Press, 2025, pp. 459–464.
11. C. H. Chiu, Y. J. Chen, Y. Wu, Y. Shi, and T. Y. Ho, “Achieve fairness without demographics for dermatological disease diagnosis,” Medical Image Analysis, vol. 95, p. 103188, 2024, doi: 10.1016/j.media.2024.103188.
12. P. Kaushik, Y. Chopra, A. Kajla, M. Poonia, A. Khan, and D. Yadav, “AI-Powered Dermatology: Achieving Dermatologist-Grade Skin Cancer Classification,” in 2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), 2024, vol. 2, pp. 1–6.

13. A. De, N. Mishra, and H. T. Chang, “An approach to the dermatological classification of histopathological skin images using a hybridized CNN-DenseNet model,” PeerJ Computer Science, vol. 10, p. e1884, 2024, doi: 10.7717/peerj-cs.1884.
14. M. S. Rashid, G. Gilanie, S. Naveed, S. Cheema, and M. Sajid, “Automated detection and classification of psoriasis types using deep neural networks from dermatology images,” Signal, Image and Video Processing, vol. 18, no. 1, pp. 163–172, 2024, doi: 10.1007/s11760-023-02722-9.
15. K. Sudharson, K. S. Essakki, V. R. Thanujaa, and S. Varsha, “Dermatec: Transformative AI-Driven Platform for Comprehensive Skin Disease Monitoring and Dermatologist Recommendations,” in 2024 IEEE International Students’ Conference on Electrical, Electronics and Computer Science, SCEECS 2024, 2024, pp. 1–5. doi: 10.1109/SCEECS61402.2024.10481923.
16. X. She, H. Lu, Q. Liu, P. Xie, and Q. Xia, “Dermatological infrared thermal imaging with human-machine interaction image diagnostics interface using DenseNet,” Journal of Radiation Research and Applied Sciences, vol. 17, no. 1, p. 100826, 2024, doi: 10.1016/j.jrras.2024.100826.
17. M. A. Farooq, W. Yao, M. Schukat, M. A. Little, and P. Corcoran, “Derm-T2IM: Harnessing Synthetic Skin Lesion Data via Stable Diffusion Models for Enhanced Skin Disease Classification using ViT and CNN,” arXiv preprint arXiv:2401.05159, 2024, doi: 10.1109/EMBC53108.2024.10781852.
18. N. Kashyap and A. K. Kashyap, “Enhanced Skin Disease Detection and Classification System Using Deep Learning Technique,” International Journal of Advanced Technology and Social Sciences, vol. 2, no. 1, pp. 93–104, 2024, doi: 10.59890/ijatss.v2i1.1292.
19. A. R. M Abbas, M Arslan, R A Bhatty, F Yousaf, A A Khan, “Enhanced Skin DiseaseDiagnosis through Convolutional Neural Networks and Data Augmentation Techniques,” Journal of Computing & Biomedical Informatics, vol. 07, no. 01, pp. 87–106, 2024.
20. B. V. S. Babu, D. T. Ram, C. Aakash, J. A. Chakravarthy, and B. L. Nancharamma, “Enhancing Dermatological Diagnosis with Optimized Deep Learning Lenet Model,” in 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems, ADICS 2024, 2024, pp. 1–6. doi: 10.1109/ADICS58448.2024.10533629.
21. I. Manole, A. I. Butacu, R. N. Bejan, and G. S. Tiplica, “Enhancing Dermatological Diagnostics with EfficientNet: A Deep Learning Approach,” Bioengineering, vol. 11, no. 8, p. 810, 2024, doi: 10.3390/bioengineering11080810.
22. G. T. Askale, D. A. Assress, A. O. Salau, and A. B. Yibel, “Ensemble of Deep CNN Models for Human Skin Disease Classification,” International Journal of Imaging Systems and Technology , vol. 34, no. 4, p. e23121, 2024, doi: 10.1002/ima.23121.
23. N. Spolaôr et al., “Fine-tuning pre-trained neural networks for medical image classification in small clinical datasets,” Multimedia Tools and Applications, vol. 83, no. 9, pp. 27305–27329, 2024, doi: 10.1007/s11042-023-16529-w.
24. Y. Nasir, K. Kadian, V. Kumar, and A. Wary, “Harnessing Quantum Computing: A Comparative Study in Skin Disease Detection with Traditional ML,” in Lecture Notes in Networks and Systems, 2024, vol. 946, pp. 361–370. doi: 10.1007/978-981-97-1323-3_30.
25. K. Vayadande, A. A. Bhosle, R. G. Pawar, D. J. Joshi, P. A. Bailke, and O. Lohade, “Innovative approaches for skin disease identification in machine learning: A comprehensive study,” Oral Oncology Reports, vol. 10, p. 100365, 2024, doi: 10.1016/j.oor.2024.100365.
26. Y. Nasir, K. Kadian, A. Sharma, and V. Dwivedi, “Interpretable machine learning for dermatological disease detection: Bridging the gap between accuracy and explainability,” Computers in Biology and Medicine, vol. 179, p. 108919, 2024, doi: 10.1016/j.compbiomed.2024.108919.
27. K. Abhishek, A. Jain, and G. Hamarneh, “Investigating the Quality of DermaMNIST and Fitzpatrick17k Dermatological Image Datasets,” Scientific Data, vol. 12, no. 1, p. 196, 2024.
28. J. Tian, K. Zhu, and W. Fan, “Mpox-PyramidTransferNet: A Hierarchical Transfer Learning Framework for Monkeypox and Dermatological Disease Classification,” in ACM International Conference Proceeding Series, 2024, pp. 404–410. doi: 10.1145/3651671.3651679.
29. S. I. Abir et al., “Precision Lesion Analysis and Classification in Dermatological Imaging through Advanced Convolutional Architectures,” Journal of Computer Science and Technology Studies, vol. 6, no. 5, pp. 168–180, 2024.
30. A. Kaur, V. Kukreja, N. Thapliyal, M. Manwal, and R. Sharma, “Pre-trained MobileNet Model to Revolutionise Dermatological Diagnostics: A Multi-Class Skin Disease Detection,” in 2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation, IATMSI 2024, 2024, vol. 2, pp. 1–5. doi: 10.1109/IATMSI60426.2024.10503199.
31. U. Nadir et al., “Protocol of a scoping review of outcome domains in dermatology,” BMJ Open, vol. 14, no. 2, p. e079632, 2024, doi: 10.1136/bmjopen-2023-079632.
32. Nirupama and Virupakshappa, “Survey on Classification of Skin Diseases Using Machine Learning Techniques,” in 2024 3rd International Conference on Power Electronics and IoT Applications in Renewable Energy and its Control, PARC 2024, 2024, pp. 135–140. doi: 10.1109/PARC59193.2024.10486701.
33. A. Pundhir, S. Verma, and B. Raman, “Towards Ethical Dermatology: Mitigating Bias in Skin Condition Classification,” in 2024 International Joint Conference on Neural Networks (IJCNN), 2024, pp. 1–8. doi: 10.1109/ijcnn60899.2024.10650487.
34. B. C. R. S. Furriel et al., “Artificial intelligence for skin cancer detection and classification for clinical environment: a systematic review,” Frontiers in Medicine, vol. 10, p. 1305954, 2023, doi: 10.3389/fmed.2023.1305954.