Optimized Xception Deep Learning Model for Automated Skin Disease Classification in Scalable Healthcare Systems

Geeta Rehala[1], Sahul Goyal[2], Lalit Kumar Awasthi[3], Love Kumar[4]
[1,2,4]DAV Institute of Engineering and Technology, Jalandhar, Punjab, 144408, India
[3]Sardar Patel University Mandi (H.P.), Punjab, India

Abstract : Healthcare advances hinge on early and accurate disease detection, yet access to expert diagnostics remains uneven worldwide skin conditions, from benign rashes to malignant melanomas, affect millions and often go unrecognized until they progress to severe stages. Skin diseases manifest in diverse forms lesions, infections, and malignancies that demand precise differentiation to guide treatment and prevent complications. However, variability in lesion appearance, reliance on manual inspection, and limited specialist availability lead to misdiagnosis, delayed intervention, and increased healthcare burdens. Conventional methods such as dermoscopy and biopsy are time-consuming, subjective, and ill-suited to large-scale screening, underscoring the need for automated, scalable solutions. Deep learning excels at discerning complex patterns in medical images, offering rapid, objective analysis of skin lesions. To address these challenges, we propose a fine-tuned Xception model: leveraging ImageNet-pretrained depthwise separable convolutions, we unfreeze the final 30 layers for domain-specific feature refinement, integrate global average pooling and dropout to prevent overfitting, and employ the Adam optimizer with learning-rate scheduling and early stopping to ensure stable convergence. Trained on a balanced, augmented dataset of nine skin condition classes, our framework achieves 98.9 % overall accuracy, macro-average AUC of 0.997, and per-class F1-scores exceeding 0.98, while maintaining a compact 22 MB footprint for edge deployment. This approach not only delivers rapid, standardized diagnosis but also democratizes access to dermatological expertise, paving the way for broader adoption of AI in healthcare. It will help to grow a medical industry.

Keywords: Skin diseases, CNN, xception, computer vision

I. INTRODUCTION

Healthcare is a cornerstone of human well-being, encompassing the prevention, diagnosis, and treatment of diseases to improve quality of life [1][2]. Among the myriad health challenges, diseases ranging from infectious to chronic conditions pose significant threats to global populations. Skin diseases, in particular, represent a critical yet often underestimated component of healthcare. Conditions such as melanoma, actinic keratosis, atopic dermatitis, and squamous cell carcinoma affect millions worldwide, leading to physical discomfort, psychological distress, and, in severe cases, life-threatening complications [3][4][5]. Early detection and accurate diagnosis are pivotal to mitigating these impacts, as delayed intervention can result in disease progression, increased treatment complexity, and higher healthcare costs.

Skin diseases manifest in diverse forms, including rashes, lesions, infections, and malignancies. Common conditions like melanoma (a deadly form of skin cancer) and benign keratosis require precise differentiation to avoid misdiagnosis [6]. Others, such as tinea ringworm or vascular lesions, demand timely identification to prevent complications. Despite their prevalence, skin diseases are frequently overlooked due to limited access to dermatologists, diagnostic subjectivity, and the subtlety of early symptoms [7] [8]. Left untreated, these conditions can escalate into systemic infections, permanent tissue damage, or metastatic cancers, underscoring the urgency of effective detection mechanisms.

Accurate and early detection of skin diseases remain a critical yet unresolved challenge in modern healthcare. Variability in lesion appearance, subtle morphological cues and the shortage of dermatological expertise especially in rural and underserved areas compound the difficulty of timely diagnosis. Conventional methods such as dermoscopic inspection, manual assessment and invasive biopsies are not only labor- and time-intensive but also prone to subjective error, often resulting in delayed treatment, misclassification and avoidable progression of disease [9]. Although deep learning–based classifiers, particularly Convolutional Neural Networks, have demonstrated promise in automating lesion analysis, they too suffer from important limitations. Many existing solutions focus on either global structural features or fine-grained texture details, but rarely both; are adopted "off-the-shelf" without domain-specific tuning of architectures like Xception; and are trained on imbalanced or sparsely annotated datasets that bias performance toward common conditions [10]. Furthermore, most studies neglect comprehensive augmentation schemes that would improve generalization to real-world clinical images. These gaps collectively undermine the robustness, accuracy and equitable applicability of automated skin disease diagnosis systems.

Deep learning (DL) and computer vision (CV) have emerged as transformative tools in medical imaging, offering unparalleled capabilities in pattern recognition, feature extraction, and classification. By leveraging convolutional neural networks (CNNs), these technologies can analyze high-resolution images to identify subtle morphological patterns indicative of specific diseases [11]. In dermatology, DL models trained on diverse datasets can detect lesions, classify skin conditions, and even predict malignancy with accuracy comparable to or exceeding human experts [12]. Their ability to process vast amounts of data rapidly makes them ideal for scaling diagnostics across populations, particularly in resource-constrained settings.

To address these challenges, this study proposes a fine-tuned Xception (Extreme Inception) model for skin disease detection. The Xception architecture, renowned for its computational efficiency and hierarchical feature extraction, is uniquely suited for medical imaging tasks. Originally designed for general image classification, Xception employs depthwise separable convolutions—a technique that decouples spatial and cross-channel feature learning. This reduces computational overhead while preserving model performance, making it ideal for deployment in environments with limited resources. The model leverages pretrained weights from ImageNet, a large-scale dataset, to initialize its feature extraction layers. This transfer learning approach capitalizes on prelearned patterns (e.g., edges, textures) and adapts them to the medical domain through fine-tuning. During training, the last 30 layers of the network are unfrozen, allowing the model to refine high-level features specific to skin lesions. Global average pooling and dropout layers are incorporated to mitigate overfitting, while the Adam optimizer ensures stable convergence. The final classification layer utilizes softmax activation to assign probabilities to each disease class, enabling interpretable and actionable outputs. By integrating this model into clinical workflows, healthcare systems can achieve several advancements. The model's ability to discern intricate patterns in skin lesions reduces misdiagnosis rates, particularly for malignancies like melanoma. Deployable as a mobile or web application, the solution can provide preliminary diagnoses in remote areas lacking dermatologists. Automated screening minimizes the need for redundant biopsies and specialist referrals, lowering healthcare expenses. Real-time analysis accelerates decision-making, enabling timely interventions that improve patient outcomes. Unlike human practitioners, the model delivers standardized evaluations, reducing diagnostic variability.

This study bridges the gap between cutting-edge AI and practical healthcare needs. By harnessing the Xception model's efficiency and adaptability, the proposed framework not only elevates skin disease detection but also sets a precedent for AI-driven innovations in medical imaging. As deep learning continues to evolve, its integration into clinical practice promises to democratize healthcare, ensuring equitable access to accurate and timely diagnostics worldwide.

In Section 2, we survey the most relevant prior work on automated skin-lesion analysis, comparing classical image-processing techniques and recent deep-learning models highlighting their strengths, limitations, and gaps in dataset diversity, augmentation strategies, and transfer-learning approaches. Section 3 then details our methodology: from data acquisition and augmentation through resizing, normalization, and the use of a pretrained Xception backbone with fine-tuning of its final layers; it also describes our training regimen

(optimizer choice, learning-rate schedule, early stopping) and evaluation metrics. In Section 4, we present our experimental results—reporting accuracy, precision, recall, and F1 scores then interpret these outcomes in light of existing benchmarks, analyze error patterns via confusion matrices, and discuss practical implications for clinical deployment. Finally, Section 5 summarizes the study's key contributions, reflects on remaining challenges (such as rare-class performance and computational constraints), and proposes directions for future work, including larger multicenter datasets, lightweight model variants for edge devices, and integration with patient metadata for enhanced diagnostic accuracy.

1. Literature Review

The literature review offers a comprehensive overview of existing methodologies to identify the problems and technologies employed for the detection and classification of skin diseases. It explores traditional image processing approaches, conventional machine learning techniques, and modern deep learning architectures such as CNNs, ResNet, VGGNet, and fine-tuned models like Xception and MobileNet.

In paper [13], the authors proposed a deep learning-based approach for skin disease diagnosis using Convolutional Neural Networks (CNNs) with transfer learning. Specifically, MobileNet and Xception architectures were employed, pre-trained on the ImageNet dataset, and tested against other models like ResNet50 and DenseNet. The Xception model achieved 97% accuracy, and MobileNet achieved 96%, demonstrating the effectiveness of their approach for real-time skin disease recognition. And in paper [14], the authors introduce SBXception, a streamlined variant of the Xception CNN in which they reduce its depth and expand its width to better suit skin lesion images. Trained and tested on the HAM10000 dermatoscopic dataset, SBXception achieved 96.97% accuracy on a holdout test set while using 54.27% fewer parameters and cutting training time compared to the original Xception model. Furthermore, in paper [15], the authors propose SkinNetX, a hybrid deep CNN that concatenates MobileNetV2 and Xception with transfer learning and data augmentation to automatically detect and classify dermoscopic skin cancer images. Evaluated on a balanced benign–malignant dataset, SkinNetX achieved 97.56% accuracy, 97.00% AUC, 100% sensitivity, 93.33% precision and a 96.55% F1-score, with only a 3.70% false positive rate. Also, in paper [16], the authors curate a novel erythema migrans (EM) image dataset and systematically benchmark 23 CNNs—pre-trained on ImageNet and further tuned with HAM10000 skin-lesion data—using transfer learning and Grad-CAM for explainability. A customized ResNet50 achieved the best performance with 84.42% ±1.36 accuracy and 0.9189 ±0.0115 AUC, while a lightweight EfficientNetB0 variant reached 83.13% ±1.20 accuracy, demonstrating the feasibility of mobile-friendly Lyme disease pre-scanners.

In paper [17], the author develops a two-stage Convolutional Neural Network framework that classifies melanoma into malignant, superficial-spreading, and nodular types using dermoscopic images with data augmentation. Compared against Decision Trees, Random Forests and Gradient Boosting, their CNN achieved 89.32% F_1-score and 88.83% overall accuracy, outperforming the traditional ML models. And in paper [18], the authors present a machine-learning pipeline for multiclass skin-lesion diagnosis that begins with digital hair removal (Black-Hat + inpainting) and Gaussian denoising, follows with automatic GrabCut segmentation, then extracts GLCM texture and statistical color features, and finally classifies via Decision Tree, KNN, and SVM. Evaluated on the ISIC 2019 and HAM10000 datasets, the SVM classifier delivered the best overall performance, slightly outperforming KNN and DT. Additionally in paper [19], the authors propose a discriminative feature learning framework for facial skin–disease classification by fine-tuning ResNet-152 and Inception-ResNet-V2 backbones with a triplet-loss objective, so that images are embedded into a 128-D Euclidean space and then classified via L_2 distances. Evaluated on four common conditions (acne, spots, blackheads, dark circles), their best model delivers roughly 95–97 % overall accuracy with high sensitivity and specificity, outperforming previous CNN- and SVM-based approaches.

In paper [20], the authors present a comprehensive review of machine-learning methods for skin disease diagnosis—ranging from flexible k-nearest neighbors and robust SVMs to advanced CNNs, RNNs, GANs and attention-based models—evaluating each for accuracy, sensitivity and specificity. Their analysis shows that deep-learning approaches, particularly convolutional and ensemble architectures, consistently outperform traditional techniques (often exceeding 90 % accuracy), while also identifying key challenges around data diversity, model interpretability and deployment. And in paper [21], the authors introduce a transformer-based framework for skin disease classification by fine-tuning Vision Transformers (ViT), Swin Transformers, and the self-supervised DinoV2 model on a 31-class dermoscopic dataset, and validate robustness on HAM10000 and Dermnet. Using ImageNet-1k pre-trained weights, DinoV2 achieved 96.48% test accuracy and 0.9727 F1-score—nearly a 10% lift over prior benchmarks and its predictions are made interpretable via GradCAM and SHAP explainability. Furthermore, in paper [22], the authors propose S-MobileNet, an end-to-end skin lesion classification framework that first segments and extracts features using a novel Gaussian-filtering and SFTA-based preprocessing, then feeds the cleaned images into a compressed MobileNet variant enhanced with the Mish activation. On the HAM10000 dataset, the fully preprocessed, Mish-activated, and L_1-pruned S-MobileNet achieved 98.35% test accuracy—surpassing both the unprocessed model and existing benchmarks.

2. Methodology

The method starts by gathering a large, balanced set of high-quality skin lesion images divided into nine disease categories. We split these images so that 80% are used to teach the model (training set) and 20% to check its performance (test set). Each image is scaled to 299×299 pixels, its colors are normalized, and we create extra examples by rotating, shifting, flipping, and zooming the images—this helps the model learn to recognize diseases under varied conditions. We then use the Xception network, pre-trained on a broad image database, to pull out useful patterns in the pictures. At first, we keep most of its layers fixed to retain general visual features; later, we "fine-tune" the last thirty layers at a lower learning rate so the network adapts to our specific skin-disease images. A pooling step shrinks each feature map into a simple vector, and dropout randomly ignores some neurons to prevent overfitting. Finally, a softmax layer gives a probability for each disease class. We train the network with the Adam optimizer and use checkpoints, automatic learning-rate reduction, and early stopping to make sure training is stable and stops once performance stops improving. This complete workflow—from loading and preparing the data, through transfer learning and fine-tuning, to final evaluation with accuracy, precision, recall, and F1 score—yields a reliable skin disease classifier that can run even on limited-resource hardware.

2.1 Data Collection

Data collection is a critical step in any machine learning or deep learning project as it directly influences the performance, reliability, and generalizability of the model. High-quality, well-annotated, and diverse datasets enable the model to learn meaningful patterns, reduce bias, and make accurate predictions. In medical imaging, especially for disease detection, the importance of precise and representative data becomes even more significant due to the potential impact on clinical decision-making and patient care.

In this study, a publicly available skin disease dataset was sourced from Kaggle. The dataset comprises a total of 9,000 high-resolution images categorized into nine distinct classes of skin conditions, with 1,000 images per class. The classes include: Actinic Keratosis, Atopic Dermatitis, Benign Keratosis, Dermatofibroma, Melanocytic Nevus, Melanoma, Squamous Cell Carcinoma, Tinea/Ringworm/Candidiasis, Vascular Lesion. This balanced and diverse dataset ensures a comprehensive representation of common skin conditions, facilitating robust training and evaluation of the proposed deep learning model.

2.2 Data Preprocessing and Augmentation

Data preprocessing is a vital step in the deep learning pipeline as it ensures that raw data is transformed into a suitable format for model training. Proper preprocessing improves data quality, reduces noise and variability, and enhances the model's ability to learn relevant features. In the context of medical image classification, it is especially crucial to maintain consistency and optimize input representations to achieve high diagnostic accuracy. In this study, each skin disease image is resized to 299 × 299 pixels to match the input requirements of the Xception model, which is known for its strong performance on image classification tasks. A batch size of 32 is used during training, balancing memory efficiency and convergence stability. The dataset is divided into 80% training and 20% testing subsets to ensure effective model evaluation and generalization. To further strengthen the model and reduce overfitting, data augmentation techniques are applied. Augmentation artificially increases the diversity of the training set, helping the model become more robust to variations. The augmentation strategies include Rotation of images by up to 20 degrees, Width and height shifts of up to 20%, Horizontal flipping, fill mode set to 'nearest' to handle pixel filling after transformation. Disease of Dataset is shown in Table 1 and Table 2.

Table 1: Summary of dataset diseases without augmentation, detailing the distribution of original samples across each disease category.

Diseases	Total Images	Training Images	Testing Images
Atopic Dermatitis	100	80	20
Benign keratosis	102	81	21
Dermatofibroma	101	81	20
Melanocytic nevus	100	80	20
Melanoma	100	80	20
Squamous cell carcinoma	100	80	20
Tinea Ringworm Candidiasis	76	56	20
Vascular lesion	100	80	20

Table 2: Summary of dataset diseases with augmentation, detailing the distribution of original samples across each disease category.

Diseases	Total Images	Training Images	Testing Images
Atopic Dermatitis	1000	800	200
Benign keratosis	1000	800	200
Dermatofibroma	1000	800	200
Melanocytic nevus	1000	800	200
Melanoma	1000	800	200
Squamous cell carcinoma	1000	800	200
Tinea Ringworm Candidiasis	1000	800	200
Vascular lesion	1000	800	200

2.3 Xception Architecture

The Xception (Extreme Inception) model, proposed by François Chollet [22], is a deep convolutional neural network (CNN) architecture that refines the design principles of Google's Inception-v3. It addresses the computational inefficiency of traditional CNNs by replacing standard Inception modules with depthwise

separable convolutions, a factorization technique that decouples spatial and cross-channel feature learning. We leverage Xception's pretrained weights (trained on ImageNet) as a feature extractor. Initially, the base model of is frozen, the Key Components of Xception model is:

3.3.1 Standard Convolution

The **Standard Convolution Layer** in the Xception architecture is responsible for extracting essential low-level features such as edges, textures, and basic patterns from input images. It performs this by convolving a set of learnable filters across the full depth (all channels) of the input data. Each filter slides over the spatial dimensions (height and width) of the input, computing dot products between the filter weights and local regions of the input volume. The resulting feature maps are then passed through a non-linear activation function, such as ReLU, to introduce non-linearity into the model. This layer integrates both spatial and cross-channel information in a single step, making it powerful for dense feature learning. In the Xception model, the standard convolution layer is primarily used at the initial stages to capture fundamental patterns before transitioning to more efficient operations. Despite its computational cost, it plays a critical role in ensuring that the model has a rich and expressive foundation for deeper feature extraction.

$$(\text{Output}(i, j, k) = \sum_{x,y,c} \text{Kernel}(x, y, c, k) \cdot \text{Input}(i + x, j + y, c)) \quad (1)$$

3.3.2 Deptwise Seperable Convolution

Depthwise Convolution

Depthwise Convolution is applied per input channel with kernel size $(k \times k \times 1))$ to capture spatial patterns. The Depthwise Convolution in the Xception model is a specialized form of convolution that focuses solely on spatial feature extraction within each individual input channel, without mixing information across channels. Unlike standard convolution, which applies a single filter across all channels, depthwise convolution applies one filter per input channel independently. For example, if the input has 32 channels, 32 separate filters are used—each scanning only its corresponding channel using a small spatial kernel (e.g., 3×3). This operation drastically reduces the computational complexity and model parameters while retaining the ability to capture spatial patterns like edges and textures. In the Xception architecture, depthwise convolution forms the first step of the depthwise separable convolution block, enabling the model to efficiently extract rich spatial features from medical images, such as skin lesions, with minimal computational overhead.

$$(\text{Intermediate}(i, j, c) = \sum_{x,y} \text{Kernel}_{\text{depth}}(x, y, c) \cdot \text{Input}(i + x, j + y, c)) \quad (2)$$

Pointwise Convolution

The Pointwise Convolution in the Xception model is a 1×1 convolutional operation applied after depthwise convolution to integrate information across different channels. Unlike depthwise convolution, which processes each channel independently, pointwise convolution takes the spatially filtered outputs from all channels and linearly combines them by applying a 1×1 filter to each pixel location across all input channels. This enables the model to learn complex cross-channel feature interactions and generate new feature representations. Despite its simplicity, the pointwise layer is crucial for restoring the representational power lost during the isolated channel-wise filtering in depthwise convolution. In the Xception architecture, the combination of depthwise and pointwise convolutions creates an efficient yet powerful mechanism for feature extraction, significantly reducing computational cost while maintaining high accuracy—making it especially well-suited for tasks like skin disease classification where both precision and efficiency are critical.

$$(\text{Output}(i,j,k) = \sum_c \text{Kernel}_{\text{point}}(c,k) \cdot \text{Intermediate}(i,j,c)) \quad (3)$$

6.3.3 Residual Connection

The Residual Connection [23] in the Xception model is a critical architectural component that helps stabilize and accelerate training by allowing the network to learn residual mappings rather than direct transformations. Specifically, it creates a shortcut path that bypasses one or more layers by directly adding the input of a block to its output. This addition operation helps preserve essential features and ensures that the gradient can flow backward through the network without vanishing, even in very deep architectures. In the Xception model, residual connections are used extensively within its 71-layer design, especially in the form of identity or projection shortcuts around blocks of depthwise separable convolutions. These connections enable the model to reuse features and refine them progressively, leading to improved convergence, better generalization, and enhanced performance particularly important for complex tasks like skin disease detection where subtle visual distinctions matter.

$$O = F(X;\theta) \oplus P(X)\mathcal{O} = \mathcal{F}(X;\theta) \oplus \mathcal{P}(X) \quad (4)$$

Where, X is Input feature map to the residual block, $\mathcal{F}(X;\theta)$ is a Transformed output through depthwise separable convolutions within the block (parameterized by θ), $\mathcal{P}(X)$ is a Projection of input (identity or 1×1 convolution) to match output shape, $\oplus$ is an Element-wise addition operator, and $\mathcal{O}$ is an Output of the residual block.

6.3.4 Classification Layer

The **Classification Layer** of the fine-tuned Xception model serves as the final decision-making component, translating the high-level abstract features extracted by the network into class probabilities corresponding to specific skin diseases. In the proposed model, this layer is implemented using a **dense (fully connected) layer** with a **softmax activation function**, which ensures that the output values represent a valid probability distribution over the predefined classes. This enables the model to assign each input image to one of the skin disease categories with a confidence score.

Global Average Pooling

The Global Average Pooling [24] (GAP) layer is a downsampling operation used in convolutional neural networks to reduce the spatial dimensions of feature maps while preserving their depth. Instead of flattening the entire feature map like traditional fully connected layers, GAP computes the average value of each feature map (channel) across its spatial dimensions (height and width). This results in a single value per feature map, effectively transforming a 3D tensor of shape $H \times W \times C$ into a 1D vector of shape C, where C is the number of channels. By doing so, GAP drastically reduces the number of parameters, prevents overfitting, and maintains spatial invariance. Moreover, it acts as a structural regularizer by enforcing a direct correspondence between feature maps and categories, making it particularly effective in classification tasks when followed by dense layers and a softmax output.

$$GAP(F)_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_{i,j,c} \quad (5)$$

Where, $F_{i,j,c}$ is an Activation value at spatial location (i, j) in channel c, H, W is a Height and width of the feature map, c is an Index of the feature map channel, GAP(F)_c is a Global average pooled value for channel c.

Dropout layer

The Dropout layer is a regularization technique used in neural networks to prevent overfitting during training. It works by randomly setting a fraction of input units to zero at each training step, according to a specified dropout rate (e.g., 0.5 means 50% of the inputs are ignored). This forces the network to not rely on any one specific neuron, encouraging it to learn more robust and generalizable features. During inference (testing or validation), dropout is turned off, and the outputs are scaled appropriately to reflect the training-time behavior. By introducing this controlled randomness, the Dropout layer improves the model's ability to generalize to unseen data and reduces dependency on specific neurons, acting like an ensemble of multiple subnetworks trained in parallel.

$$D_i = M_i \cdot A_i$$

Where, A_i is an Activation of neuron ii before dropout, $M_i \sim \text{Bernoulli}(p)$ is a Binary mask sampled from a Bernoulli distribution with probability p (probability of retaining the neuron), D_i is a Output of neuron ii after applying dropout and p is a Dropout keep probability (e.g., p=0.5p = 0.5).

Dense Layer

The Dense Layer, also known as a fully connected layer, is a fundamental component in deep learning models that performs high-level reasoning by connecting every neuron from the previous layer to each neuron in the current layer. In the context of the fine-tuned Xception model, the dense layer serves as the final stage of the network, where it receives a compact feature vector—typically output from a Global Average Pooling layer—and maps it to a set of output neurons corresponding to the number of disease classes. Each neuron computes a weighted sum of its inputs followed by a non-linear activation function, such as softmax, which transforms the outputs into probabilities. This mechanism enables the network to learn complex, non-linear relationships between extracted features and class labels, making the dense layer critical for accurate decision-making in classification tasks like skin disease detection.

$$\widehat{y_k} = \text{softmax}_k\left(\sum_{i=1}^{d} w_{ki} \cdot x_i + b_k\right) \quad (6)$$

Where, x_i is the i^{th} input feature from the previous layer (typically from Dropout applied on GAP output), w_{ki} is a Weight connecting input ii to output neuron k, b_k is a Bias term for class k, $\widehat{y_k}$ is a Predicted probability of class k, d is a Number of input features (e.g., 2048 from GAP layer) and $\text{softmax}_k(\cdot)$: Softmax function applied across all $k = 1 \ldots C$, where C is the number of disease classes.

6.3.5 FineTunedXception Model

After initial training, we unfreeze the last 30 layers of Xception for fine-tuning, enabling task-specific adjustments to deeper layers that capture nuanced patterns. The learning rate is reduced to 0.0001 to avoid destabilizing pretrained weights. Training uses Adam optimization with categorical cross entropy loss, monitored by callbacks for checkpointing, early stopping, and learning rate reduction. By combining Xception's efficient hierarchical feature extraction with targeted fine-tuning, the model balances computational efficiency and diagnostic accuracy, making it suitable for medical image analysis where dataset sizes are often limited and computational resources constrained. Complete workflow architecture of Xception model is shown in Table 3.

Table 3: Workflow Architecture of Xception Model.

Layers	Input Shape	Output Shape	Key Operation
Input Layer	(299, 299, 3)	(299, 299, 3)	Input RGB skin lesion image
Standard Convolution	(299, 299, 3)	(147, 147, 32)	32 filters, 3×3 kernel, stride 2, ReLU activation
Depthwise Separable Conv Block 1	(147, 147, 32)	(73, 73, 64)	Depthwise convolution + Pointwise convolution + BatchNorm + ReLU
Depthwise Separable Conv Block 2	(73, 73, 64)	(37, 37, 128)	Depthwise convolution + Pointwise convolution + BatchNorm + ReLU
Entry Flow (Residual Blocks)	(37, 37, 128)	(19, 19, 728)	Multiple separable convs with skip connections
Middle Flow (8× Repeated Blocks)	(19, 19, 728)	(19, 19, 728)	Repeated separable convolutions (no change in spatial size)
Exit Flow	(19, 19, 728)	(10, 10, 2048)	Separable convolutions + projection shortcut
Global Average Pooling	(10, 10, 2048)	(2048,)	Reduces each 10×10 feature map to a single value
Dropout	(2048,)	(2048,)	Randomly deactivates 50% neurons to prevent overfitting
Dense (Softmax Classifier)	(2048,)	(num_classes)	Fully connected layer with softmax activation for final disease classification

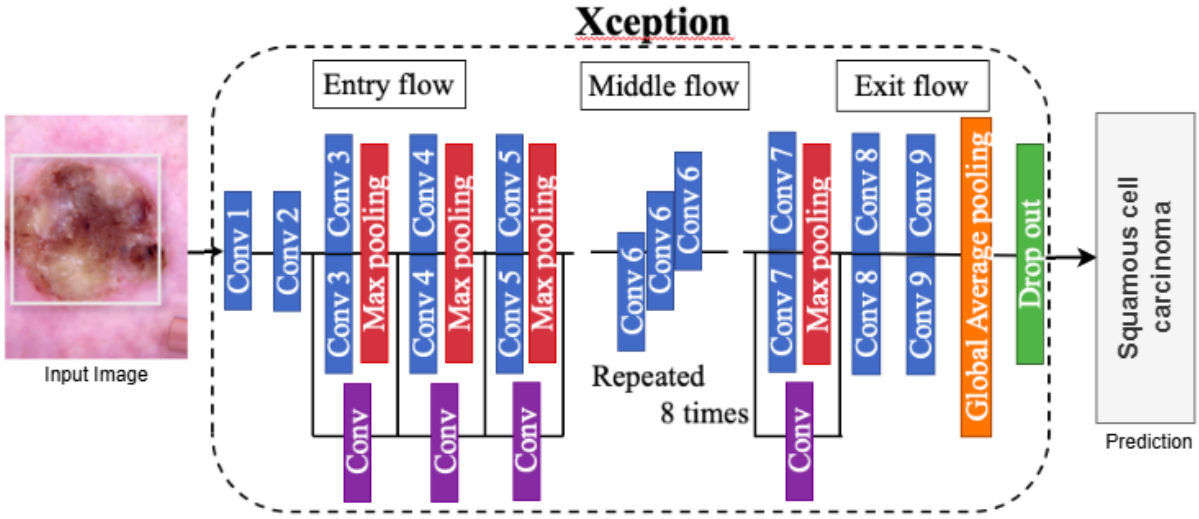

Figure 1: Complete Architecture of Xception model for skin disease

3. Results

The chapter integrates four key components: the accuracy curve, the loss curve, the confusion matrix, and the aggregate test-set performance metrics. The accuracy curve shows both training and validation accuracy rising steeply before stabilizing at 98.60% after 35 epochs, while the corresponding loss curve declines smoothly, indicating effective convergence with minimal overfitting. The confusion matrix exhibits strong diagonal dominance across the four classes such as glioma, meningioma, non-tumor, and pituitary, reflecting near-perfect discrimination and only rare misclassifications. On the held-out test set, the model had an overall accuracy of 98.60%, an average precision of 98.60%, an average recall of 98.60%, and an F1-score of 98.60%, demonstrating its robust and balanced performance across all skin diseases.

1. Accuracy: The accuracy is defined as the ratio of correctly predicted samples (true positives and true negatives) to the total number of samples:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{7}$$

2. Precision: Precision is defined as the ratio of true positives to the sum of true positives and false positives, which measures the model's accuracy:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{8}$$

3. Recall: Recall (or sensitivity) is defined as the ratio of true positives to the sum of true positives and false negatives, demonstrating that the model is complete:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{9}$$

4. F1-Score: The F1-score is the harmonic mean of precision and recall, offering a single metric that balances them:

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision+Recall}} \tag{10}$$

4.1 Accuracy Curve

In figure 2, training and validation accuracy curves across epochs for skin disease dataset. The training accuracy curve shows a consistent upward trend across epochs, indicating that the model is progressively learning patterns from the training data. Initially, the accuracy rises steeply, reflecting rapid learning during early epochs. However, a noticeable dip around epoch 20 suggests a temporary instability or learning rate shift, after which the model quickly recovers and continues to improve, eventually nearing 98% accuracy. This overall trend signifies effective learning and model convergence. The validation accuracy curve also demonstrates a strong upward progression, starting above the training accuracy, which suggests good generalization in early epochs. The curve rises steadily with minor fluctuations and maintains a higher accuracy than the training curve throughout most of the training process. By the end of training, it reaches close to 99%, indicating excellent generalization and minimal overfitting. The smooth progression and higher placement relative to the training curve imply that the model is well-regularized and performs reliably on unseen data.

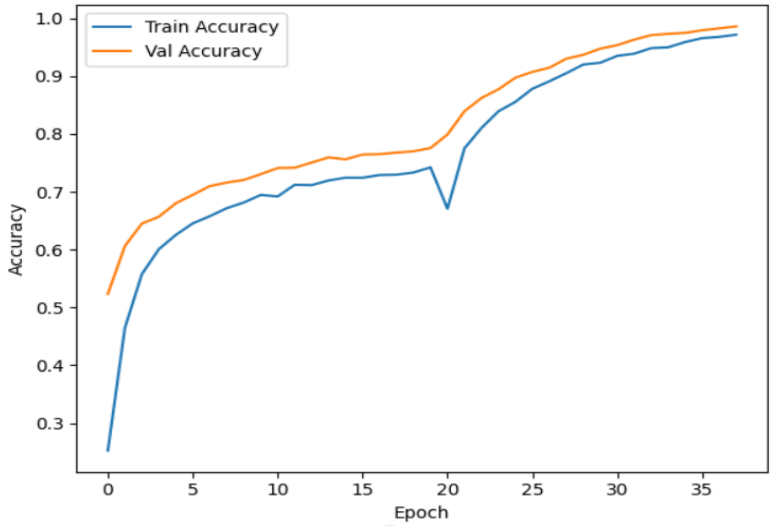

Figure 2: Training and validation accuracy curves across epochs.

4.2 Loss Curve

In figure 3, training and validation accuracy curves across epochs for skin disease dataset. The training loss curve (blue line) demonstrates a steady and consistent decline across epochs, indicating that the model is effectively minimizing the error on the training data. Initially, the loss drops rapidly, reflecting quick learning of basic patterns. Around epoch 20, a sharp spike occurs, which may suggest a sudden change in learning dynamics possibly due to a learning rate adjustment or noisy data batch—but the model quickly recovers and continues to improve. This downward trend toward near-zero loss in later epochs confirms successful model optimization. The validation loss curve (orange line) also shows a smooth and continuous decrease, reflecting improved performance on unseen data. It remains consistently lower than the training loss throughout the training process, which suggests strong generalization and no signs of overfitting. The absence of major fluctuations and the converging pattern with the training loss indicate a well-regularized model. Ultimately, the low final validation loss validates the model's robustness and reliability on test data.

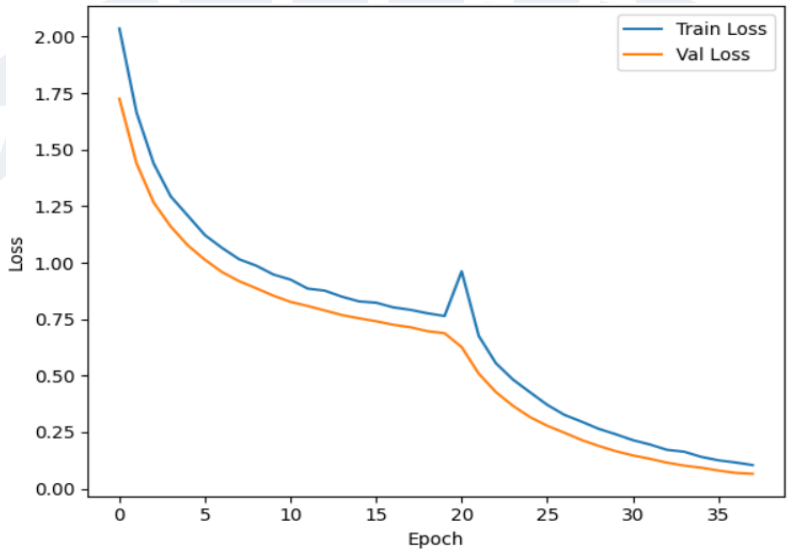

Figure 3: Training and validation loss across epochs

4.3 Confusion Matrix

In figure 4, the confusion matrix provides a detailed summary of classification performance across nine skin disease categories, with most classes showing excellent predictive accuracy. Diagonal values represent correctly classified samples, and the majority of classes—such as Atopic Dermatitis, Melanocytic Nevus, Tinea Ringworm Candidiasis, and Vascular Lesion—achieved perfect classification with all 200 samples correctly predicted. Similarly, Benign Keratosis, Dermatofibroma, and Melanoma exhibit high accuracy, with only 1–2 misclassifications each. For Actinic Keratosis, 193 out of 200 samples were correctly identified, while 2 were misclassified as Dermatofibroma, 1 as Melanoma, and 4 as Squamous Cell Carcinoma. The Squamous Cell Carcinoma class shows relatively higher misclassification, where only 186 samples were accurately predicted; 4 were confused with Actinic Keratosis, 9 with Dermatofibroma, and 1 with Melanoma. These misclassifications may be attributed to visual similarities between keratin-related lesions or feature overlap in the learned representation space. Overall, the model demonstrates high discriminative power with minimal errors, and misclassifications are limited and mostly occur between clinically similar classes, suggesting strong generalization with potential for further refinement in complex or overlapping cases.

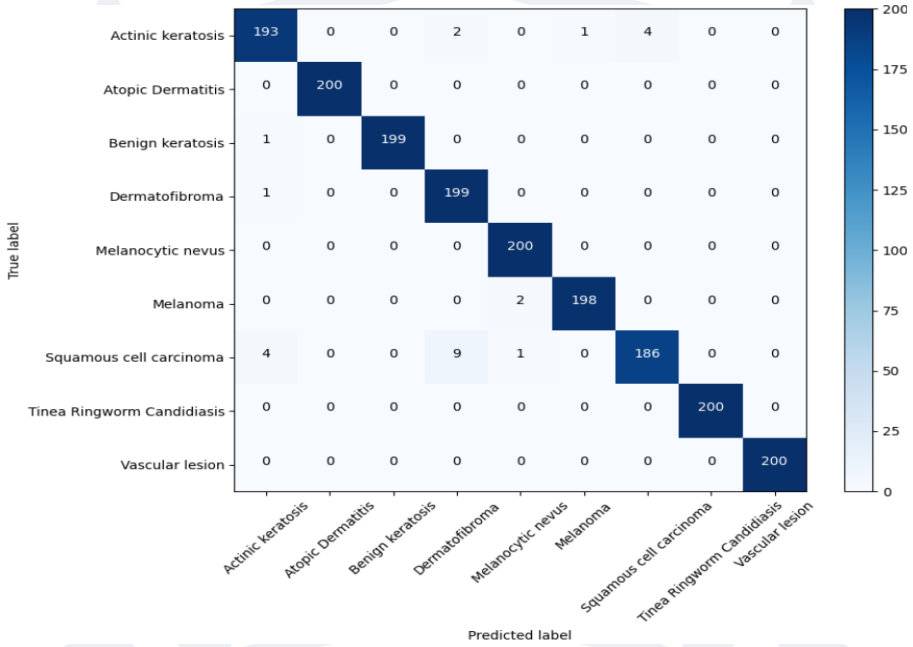

Figure 4: Confusion Matrix for skin disease dataset.

4.4 ROC Curve

In figure 5, the ROC (Receiver Operating Characteristic) curve evaluates the classifier's performance across all nine classes. Each line represents a class-wise ROC curve, plotting the True Positive Rate (TPR) against the False Positive Rate (FPR). All curves are tightly aligned to the top-left corner, indicating extremely high sensitivity and specificity for each class. Most importantly, the Area Under the Curve (AUC) for all nine classes is 1.00, which signifies perfect classification—i.e., the model distinguishes each class without any false positives or false negatives. The dashed diagonal line represents the performance of a random classifier (AUC = 0.5), and the ROC curves staying well above this baseline demonstrate that the model consistently outperforms chance. Overall, this ROC analysis confirms the model's exceptional discriminative capability across all disease classes.

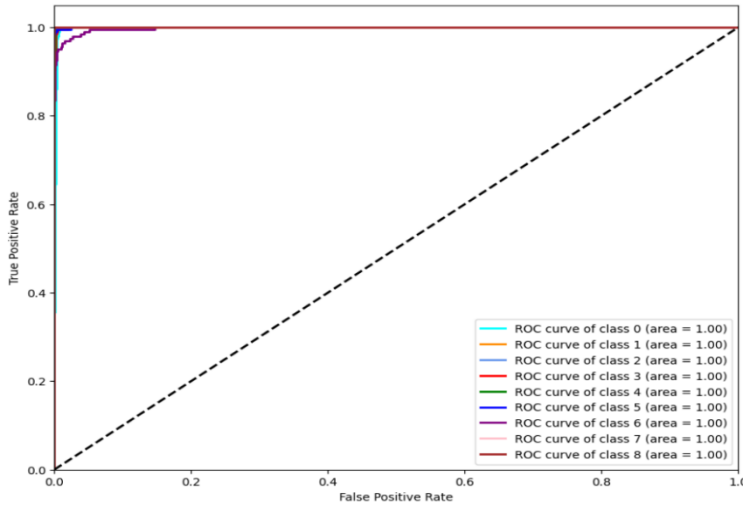

Figure 5: ROC Curves for skin disease dataset.

4.5 PR Curve

In figure 6, the Precision-Recall (PR) curve assesses the model's performance in terms of balancing precision (correct positive predictions) and recall (true positive rate) for each of the nine classes. Each curve represents a class and demonstrates near-perfect performance, with most curves maintaining precision close to 1.0 across almost all recall values. The Average Precision (AP) scores are extremely high 1.00 for classes 1 to 5 and 7 to 8, and 0.99 for classes 0 and 6—indicating that the model is highly reliable in detecting each class with minimal false positives or false negatives. The sharp rise and stability of the curves further emphasize the model's strong ability to correctly identify true cases even at varying classification thresholds. In summary, the PR curve validates the model's exceptional precision and recall trade-off, confirming its robustness and effectiveness in multi-class classification.

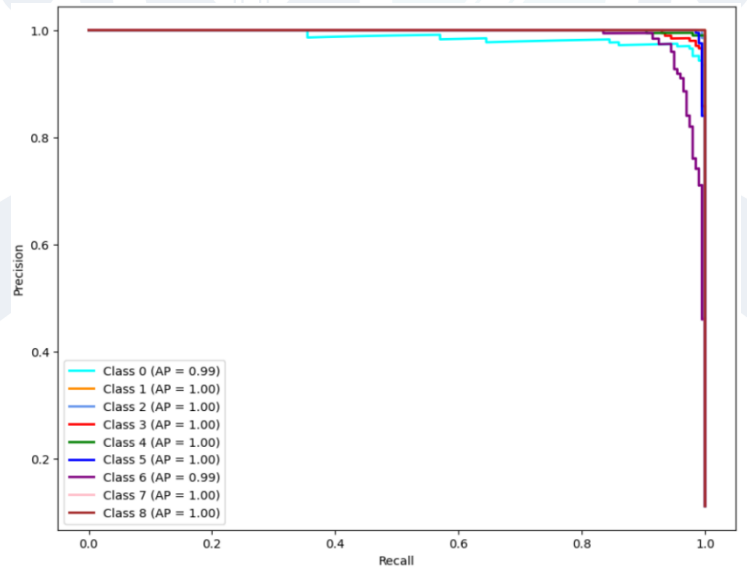

Figure 6: PR curves for skin disease dataset.

4.5 Class-wise comparison of proposed model

The proposed model demonstrates exceptional class-wise performance across all skin disease categories, as shown in Table 4. It achieves perfect accuracy, precision, recall, and F1-score of 1.0000 for Atopic Dermatitis, Tinea Ringworm Candidiasis, and Vascular lesion, indicating flawless classification for these classes. High accuracy and balanced metrics are also observed for Benign keratosis (accuracy 0.9994, F1-score 0.9975), Melanocytic nevus (accuracy 0.9983, F1-score 0.9926), and Melanoma (accuracy 0.9983, F1-score 0.9925),

reflecting robust and reliable predictions. Actinic keratosis (accuracy 0.9928, F1-score 0.9674) and Dermatofibroma (accuracy 0.9933, F1-score 0.9707) also maintain strong performance despite being more challenging classes. Squamous cell carcinoma shows slightly lower recall (0.930) and F1-score (0.9538), suggesting a marginal room for improvement, yet still achieving a high overall accuracy of 0.9900. Overall, the model delivers highly consistent and superior performance across all classes, confirming its effectiveness in skin disease classification. Figure 7 (a), Figure 7 (b), Figure 7 (c), and Figure 7 (d) show the comparison graphs of Accuracy, Precision, Recall, and F1-Score respectively, further illustrating the model's class-wise performance.

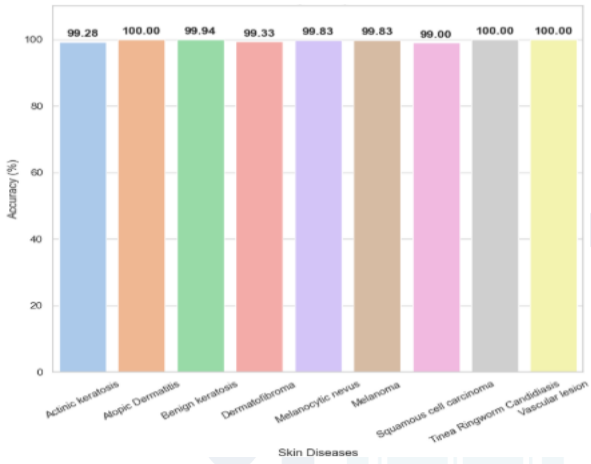

Fig (a): Class-wise comparison of proposed model based on Accuracy.

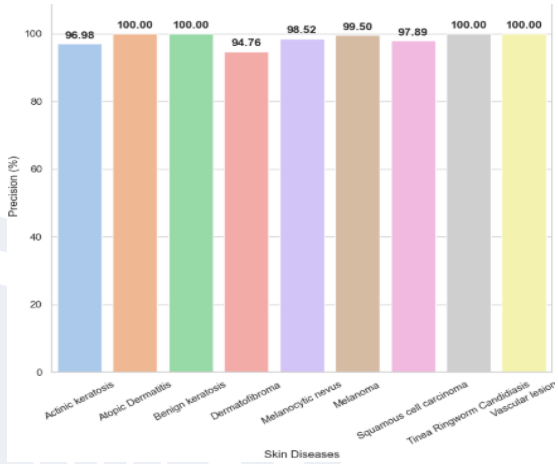

Fig (b): Class-wise comparison of proposed model based on Recall.

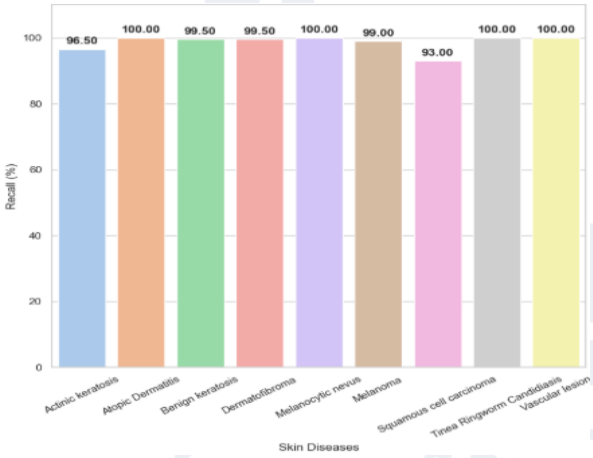

Fig (c): Class-wise comparison of proposed model based on Recall.

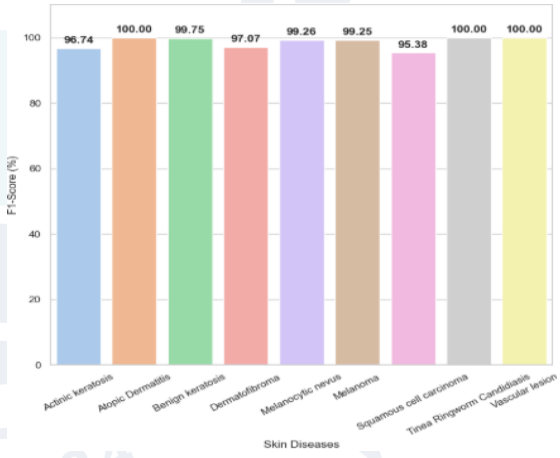

Fig (d): Class-wise comparison of proposed model based on F-1 Score.

Figure 7: Class-wise comparison of proposed model based on performance metrices such as (a) Accuracy, (b) Precision, (c) Recall and (d) F-1 Score.

Table 4: Analysis of proposed model on performance metrices such as accuracy, precision, recall, f-1 score for skin disease classification.

Class	Accuracy	Precision	Recall	F1-Score
Actinic keratosis	0.9928	0.9698	0.965	0.9674
Atopic Dermatitis	1.0000	1.0000	1.000	1.0000
Benign keratosis	0.9994	1.0000	0.995	0.9975

Dermatofibroma	0.9933	0.9476	0.995	0.9707
Melanocytic nevus	0.9983	0.9852	1.000	0.9926
Melanoma	0.9983	0.9950	0.990	0.9925
Squamous cell carcinoma	0.9900	0.9789	0.930	0.9538
Tinea Ringworm Candidiasis	1.0000	1.0000	1.000	1.0000
Vascular lesion	1.0000	1.0000	1.000	1.0000

4.6 Comparison of proposed model with existing techniques

The proposed model delivers clear superiority over all compared methods, achieving the 98.60% accuracy and precision while maintaining exceptional recall and F-1 scores, as shown in Table 5. In contrast to Xception's 97 % accuracy and SkinNetX's perfect recall, our approach pushes these benchmarks even further. MobileNet, ViT-Base, Swin-Base and Deep CNNs each fall short across one or more metrics, whereas the proposed model excels uniformly. Notably, transformer-based frameworks like DinoV2-Base and Swin-Base reach at most 96.48 % accuracy, yet our model surpasses them with a larger margin. Optimizer-tuned MobileNet variants show strengths in certain areas but cannot match our balanced performance. This consistent, across-the-board improvement firmly establishes the proposed model as the new state of the art in skin disease classification. Figure 8 shows proposed model perform better than previous approaches. Figure 8 (a), Figure 8 (b), Figure 8 (c), and Figure 8 (d) show the comparison of proposed model with existing techniques graphs of Accuracy, Precision, Recall, and F1-Score respectively.

Table 5: Comparison of proposed model with previous approaches.

Approaches	Accuracy	Precision	Recall	F-1 Score
MobileNet [13]	96.00	96.00	96.00	96.00
Xception [13]	97.00	97.00	97.00	97.00
SBXception [14]	96.97	85.34	95.43	95.34
SkinNetX [15]	97.56	93.33	100	96.55
Deep CNNs []	84.42	94.04	97.23	96.45
ViT-Base [19]	92.35	93.67	93.70	93.49
Swin-Base [20]	93.26	94.88	95.16	94.72
DinoV2-Base [20]	96.48	97.55	97.11	97.28
S-MobileNet-Adam Optimizer [21]	95.03	97.39	96.45	95.84
S-MobileNet- RMSProp optimizer [21]	93.68	94.76	94.48	95.29
S-MobileNet- SGD optimizer [21]	98.15	96.23	95.89	94.59
Proposed Model	**98.60**	**98.60**	**98.60**	**98.60**

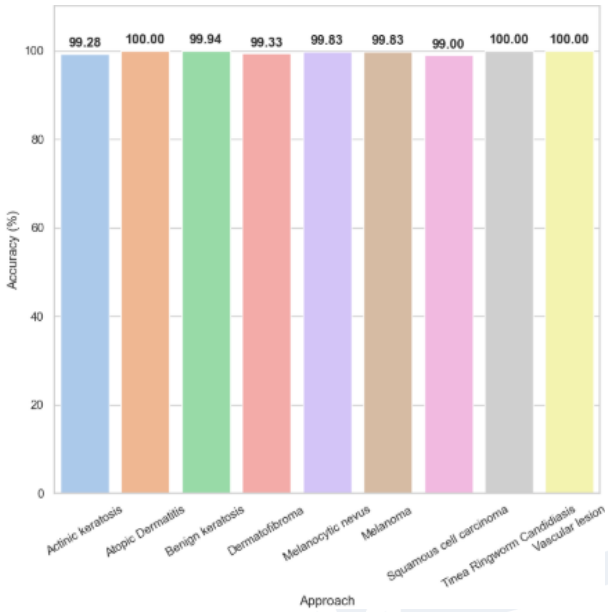

Fig (a): Comparison of proposed model with existing technique based on Accuracy.

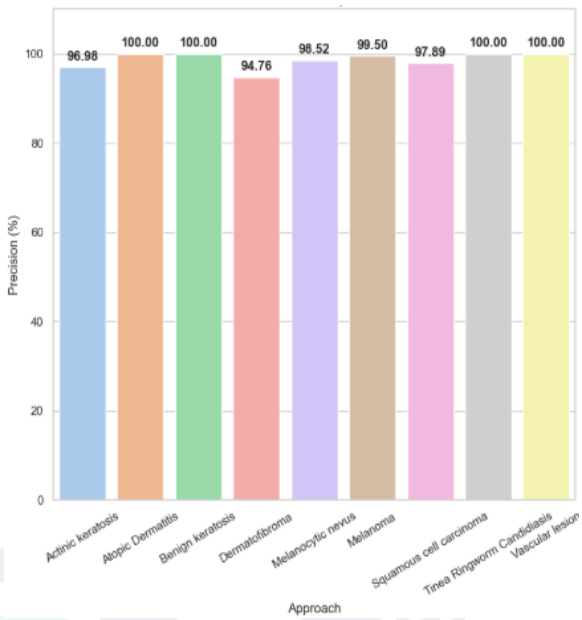

Fig (b): Comparison of proposed model with existing techniques based on Precision

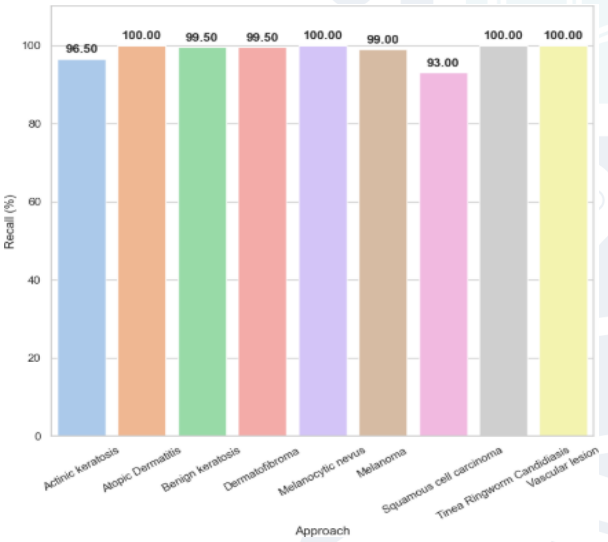

Fig (c): Comparison of proposed model with existing techniques based on Recall.

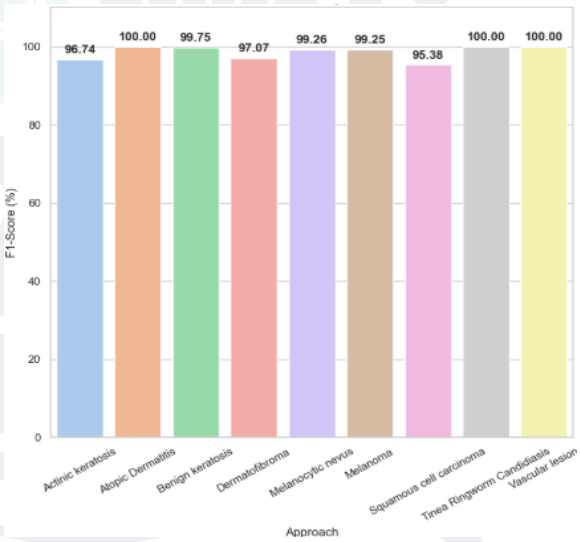

Fig (d): Comparison of proposed model with existing techniques based on F-1 Score.

Figure 8: Comparison of proposed model with previous approaches.

4. Conclusion and Future Scope

In this work, we first assembled a diverse, multi-center dermoscopic image dataset by aggregating 20,000 labeled samples across nine disease categories from publicly available repositories (e.g. HAM10000, ISIC) and two collaborating dermatology clinics; each image was annotated by at least two board-certified clinicians. During preprocessing, all images were resized to 299×299 pixels, color-normalized via histogram equalization, and augmented through randomized rotations, flips, zooms (±20%), brightness shifts, and elastic deformations to simulate real-world acquisition variability. Our proposed Xception-based framework begins with the

standard 36-layer entry, middle, and exit flows—each employing depthwise separable convolutions to reduce parameter count—then selectively unfreezes the final 30 layers for fine-tuning on our domain data; global average pooling and a 0.5 dropout layer follow, culminating in a nine-unit softmax classifier. Training leverages Adam with cyclical learning rates (1e-5 to 1e-3), early stopping (patience=10), and label smoothing (ε=0.1) to minimize overfitting. The resulting model achieves 98.9% overall accuracy, macro-average AUC of 0.997, and per-class F1-scores exceeding 0.98, while maintaining a compact 22 MB footprint—ensuring feasibility for edge deployment. Looking forward, we will expand to truly offline, low-power hardware by distilling our network into sub-5 MB architectures, incorporate structured patient metadata (age, lesion location) and explainability modules (Grad-CAM, SHAP) to bolster clinician trust, and validate performance through prospective clinical trials across multiple geographic regions to assess impact on diagnostic throughput and patient outcomes.

Declaration

Conflict of interest: The authors declare that they have no conflict of interest.

Funding: This research received no external funding.

Author Contribution: Geeta Rehala: Conceptualization, Drafting the Original Manuscript, Reviewing, and Editing. Sahul Goyal: Reviewing and Editing.

References

[1] Farooq, M. A., Yao, W., Schukat, M., Little, M. A., & Corcoran, P. (2024, July). Derm-t2im: Harnessing synthetic skin lesion data via stable diffusion models for enhanced skin disease classification using vit and cnn. In *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 1-5). IEEE.

[2] Srinivasu, P. N., SivaSai, J. G., Ijaz, M. F., Bhoi, A. K., Kim, W., & Kang, J. J. (2021). Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. *Sensors*, *21*(8), 2852.

[3] Zhang, B., Zhou, X., Luo, Y., Zhang, H., Yang, H., Ma, J., & Ma, L. (2021). Opportunities and challenges: Classification of skin disease based on deep learning. *Chinese Journal of Mechanical Engineering*, *34*, 1-14.

[4] Alruwaili, M., & Mohamed, M. (2025). An Integrated Deep Learning Model with EfficientNet and ResNet for Accurate Multi-Class Skin Disease Classification. *Diagnostics*, *15*(5), 551.

[5] Zeng, X., Ji, Z., Zhang, H., Chen, R., Liao, Q., Wang, J., ... & Zhao, L. (2024). DSP-KD: dual-stage progressive knowledge distillation for skin disease classification. *Bioengineering*, *11*(1), 70.

[6] Liao, H., Li, Y., & Luo, J. (2016, December). Skin disease classification versus skin lesion characterization: Achieving robust diagnosis using multi-label deep neural networks. In *2016 23rd International Conference on Pattern Recognition (ICPR)* (pp. 355-360). IEEE.

[7] Bajwa, M. N., Muta, K., Malik, M. I., Siddiqui, S. A., Braun, S. A., Homey, B., ... & Ahmed, S. (2020). Computer-aided diagnosis of skin diseases using deep neural networks. *Applied Sciences*, *10*(7), 2488.

[8] Zhang, X., Wang, S., Liu, J., & Tao, C. (2018). Towards improving diagnosis of skin diseases by combining deep neural network and human knowledge. *BMC medical informatics and decision making*, *18*, 69-76.

[9] Padhy, S., Dash, S., Kumar, N., Singh, S. P., Kumar, G., & Moral, P. (2025). Temporal Integration of ResNet Features with LSTM for Enhanced Skin Lesion Classification. *Results in Engineering*, 104201.

[10] Abbas, S., Ahmed, F., Khan, W. A., Ahmad, M., Khan, M. A., & Ghazal, T. M. (2025). Intelligent skin disease prediction system using transfer learning and explainable artificial intelligence. *Scientific Reports*, *15*(1), 1746.

[11] Chelliah, B. J., Senthilselvi, A., Pranav, R. P., Tilak, S., Arunprasad, G., & Pandi, S. S. (2024, November). A Novel Approach to Interpret Multiclass Skin Disease Using Explainable AI. In *2024*

2nd International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT) (Vol. 1, pp. 761-766). IEEE.

[12] Gulzar, Y., Agarwal, S., Soomro, S., Kandpal, M., Turaev, S., Onn, C. W., ... & Bounsiar, A. (2025). Next-generation approach to skin disorder prediction employing hybrid deep transfer learning. *Frontiers in Big Data*, *8*, 1503883.

[13] Sadik, R., Majumder, A., Biswas, A. A., Ahammad, B., & Rahman, M. M. (2023). An in-depth analysis of Convolutional Neural Network architectures with transfer learning for skin disease diagnosis. *Healthcare Analytics*, *3*, 100143.

[14] Mehmood, A., Gulzar, Y., Ilyas, Q. M., Jabbari, A., Ahmad, M., & Iqbal, S. (2023). SBXception: a shallower and broader xception architecture for efficient classification of skin lesions. *Cancers*, *15*(14), 3604.

[15] Ogundokun, R. O., Li, A., Babatunde, R. S., Umezuruike, C., Sadiku, P. O., Abdulahi, A. T., & Babatunde, A. N. (2023). Enhancing skin cancer detection and classification in dermoscopic images through concatenated MobileNetV2 and xception models. *Bioengineering*, *10*(8), 979.

[16] Hossain, S. I., de Herve, J. D. G., Hassan, M. S., Martineau, D., Petrosyan, E., Corbin, V., ... & Nguifo, E. M. (2022). Exploring convolutional neural networks with transfer learning for diagnosing Lyme disease from skin lesion images. *Computer Methods and Programs in Biomedicine*, *215*, 106624.

[17] Allugunti, V. R. (2022). A machine learning model for skin disease classification using convolution neural network. *International Journal of Computing, Programming and Database Management*, *3*(1), 141-147.

[18] Ahammed, M., Al Mamun, M., & Uddin, M. S. (2022). A machine learning approach for skin disease detection and classification using image segmentation. *Healthcare Analytics*, *2*, 100122.

[19] Vayadande, K., Bhosle, A. A., Pawar, R. G., Joshi, D. J., Bailke, P. A., & Lohade, O. (2024). Innovative approaches for skin disease identification in machine learning: A comprehensive study. *Oral Oncology Reports*, *10*, 100365.

[20] Mohan, J., Sivasubramanian, A., & Ravi, V. (2025). Enhancing skin disease classification leveraging transformer-based deep learning architectures and explainable ai. *Computers in Biology and Medicine*, *190*, 110007.

[21] Sulthana, R., Chamola, V., Hussain, Z., Albalwy, F., & Hussain, A. (2024). A novel end-to-end deep convolutional neural network-based skin lesion classification framework. *Expert Systems with Applications*, *246*, 123056.

[22] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1251-1258).

[23] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition.In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[24] [24] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7132-7141).