

# Convolutional Neural Network-Based Classification of Skin Lesions Using Dermoscopic Images

Sujatha Krishna<sup>1</sup>, Osamah Ibrahim Khalaf<sup>2</sup>

<sup>1</sup>Department of Computing and Information Sciences, College of Computing and Information Sciences, University of Technology and Applied Sciences, Al-Aqr, Shinas, 324, Oman

<sup>2</sup>Al-Nahrain University, Al-Nahrain Renewable Energy Research Center, Baghdad, Iraq

## Article Info

### Article history:

Received November 20, 2025

Revised January 28, 2026

Accepted February 10, 2026

### Keywords:

Skin Disease Detection

Dermoscopic Images

Deep Learning

Convolutional Neural Networks

Streamlit Deployment

## ABSTRACT

Skin cancer is one of the most common and deadliest diseases globally, in which early detection may help improve patient survival rates. In this paper, an automatic skin cancer classification framework based on deep learning using dermoscopic images is presented. Various convolutional neural network (CNN) models were trained and tested on an annotated skin lesion dataset, including MobileNet, EfficientNetB0, VGG16, ConvNet, and ResNet50. Metrics of the models were calculated by using Micro-averaged metrics to assess the general effectiveness for all the classes. ResNet50 obtained the best performance against all tested models with a micro-average accuracy of 97.75%, precision of 97.79%, recall of 97.75%, and F1 score of 97.76%. Our results indicate that the model enables accurate, consistent, and balanced classification of different skin lesion categories, including actinic keratoses, basal cell carcinoma, benign Keratosis-like lesions, dermatofibroma, and melanoma. For real-world utilization, the top performer, the ResNet50 model, was implemented in a Streamlit-based web application, which is designed to automatically predict skin diseases in dermoscopic images that were uploaded. Experimental results show that deep residual learning is effective for improving the classification performance of skin lesions, and it can become an assistive decision-making tool for dermatologists in early diagnosis and clinics.

*This is an open access article under the [CC BY-SA](#) license.*



**Corresponding Author:** Sujatha Krishna (e-mail: [Sujatha.Krishna@utas.edu.om](mailto:Sujatha.Krishna@utas.edu.om))

## 1. INTRODUCTION

Cutaneous diseases are the most frequent health problems found around the world and occur in all ages. Although most skin conditions are not life-threatening, it is important for them to be treated and diagnosed early in the development of a disease such as cancer. The worldwide increase of skin cancer has been the subject of numerous studies in medical literature, which underlines the importance of accurate and early diagnosis systems [1]. Early detection is critical to decrease disease progression and increasing survival, especially in malignant types like melanomas. Clinicophysical examination, Magnetic resonance spectroscopy, and non-invasive imaging techniques can assist dermatologists in the diagnosis of cutaneous lesions by enhancing the visualization of subsurface structures of the skin. While dermoscopy enhances diagnostic performance, interpreting dermoscopic images is skill-dependent and demands experience. It is not readily possible to manually diagnose, because of the visual resemblance between various categories of lesions, which implies the need for automatic and Computer-Aided Diagnosis (CAD) systems [2]. Automated systems strive for consistent and objective assessment to assist clinicians in reliable decision-making.

The developments of artificial intelligence and deep learning have greatly revolutionized medical image analysis. Deep learning methods, especially neural network-based architectures, have shown powerful learning capacity in detecting and classifying subtle or intricate disease patterns in medical images [3]. These models automatically extract hierarchical features from data, avoiding the cumbersome feature engineering and yielding superior classification performance. These features render deep learning well-suited for the

analysis of dermoscopic images and the detection of skin diseases. CNNs are currently the most successful deep learning method for image-based medical diagnosis. CNN models can learn and represent the discriminative spatial patterns, as well as subtle differences of medical images, automatically. It has been proven that the classification performance improves a lot by combining deep learning and efficient feature extraction for skin lesion detection [4]. Thus, CNN architectures have emerged as the cornerstone of CADs in dermatology.

Several deep learning models have been developed for automated skin lesion classification, such as combination enhancement and probabilistic neural networks. These methods are to enhance the accuracy of classification by integrating feature learning and optimization techniques [5]. Furthermore, optimized CNN-based models have also been developed to further improve the diagnostic performance of CAD systems by training network parameters and learning efficiency [6]. These enhancements enable the models to leverage complex visual patterns in dermoscopic images. Deep learning (DL)-based computer-aided diagnosis (CAD) in dermatology has demonstrated promising results for the assembly of an automated skin lesions detection and classification system to facilitate dermatologists. These approaches enable the mitigation of variability in diagnostics and enhance the effectiveness of clinical workflow [7]. Moreover, contemporary design of CNN architectures such as EfficientNet has been proven efficient in multi-class classification for skin-lesions, hence making it possible for enhanced detection between various disease types [8].

Furthermore, the role of artificial intelligence in healthcare systems has broadened to include secure and intelligent medical frameworks. The emerging AI-based health care models, such as decentralized and intelligent systems, emphasize the increasing demand for automatic medical decision-support technologies [9]. These advances showcase the opportunity that AI has to improve healthcare through automation and data-guided diagnostics. Recent research has also concentrated on refining CNN-based skin lesion classification using architectural tuning and checkpoint learning tactics. These techniques are designed to increase model generalization and strengthen discriminative capacity on dermoscopic image data from different sources [10]. Optimization methods have been used extensively in CNN architectures to enhance the training effectiveness and classification accuracy of automated skin cancer detection systems [11].

Driven by the rapid development of deep learning, automatic skin disease detection systems have been widely deployed in smart healthcare systems. Recent literature illustrates the promising ability of these convnets to perform well in large-scale skin cancer screening and intelligent diagnostic assistance within digital health platforms [12]. These advances underscore the increasing importance of AI-powered tools for improving early detection, as well as diagnostic consistency. A review was conducted on a deep learning (DL) framework for multiclass classification of dermoscopic images. The study investigates several convolutional neural network models and evaluates them in automatic skin disease identification. Dealing with systematic model construction, efficient preprocessing, and accurate evaluation is emphasized. The aim of this study is to extend the knowledge and searchable evidence for computer-aided dermatological diagnosis with an automatic and reproducible method for analyzing dermoscopic images.

The main contributions of the proposed model are:

- In this paper, a novel deep learning-based framework that facilitates automatic multi-class skin lesions classification using dermoscopic images in order to have trustworthy categorization of different skin diseases is proposed.
- Several configurations of CNNs are systematically applied and compared to test their power for disease-based skin detection and classification.
- A model architecture with an enhanced deep CNN structure is proposed to enhance feature extraction and classification quality of complex dermoscopic skin lesion images.
- We follow a comprehensive evaluation scheme that is based on micro-averaging accuracy, precision, recall, and F1-score to maintain balance and consistency of the performance across all lesion categories.
- The proposed system, developed as an interactive Streamlit web application enabling live skin disease prediction from uploaded dermoscopic images, is a proof of concept for practical applicability.
- The automated, rapid, and stable approach will contribute to the development of a computer-aided diagnosis system in dermatology for early identification of skin disease.

The rest of this paper is arranged as follows. The literature review is given in Section 2. In Section 3, we present the dataset and methodology. Section 4 describes the results and evaluation. Section 5 concludes with future work.

## 2. LITERATURE REVIEW

Recent developments in AI and medical image analysis have brought about great achievements in automatic disease detection and diagnosis. There is a growing volume of research on deep learning, machine learning, and optimization models for analyzing medical and dermoscopic images to assist in the early diagnosis of diseases. Computer-aided diagnosis systems have shown a strong potential to improve classification performance, reduce diagnostic variability, and help healthcare practitioners in clinical decision making. This section summarizes the related works about automatic skin disease recognition, with emphasis on methodologies, contributions, and weaknesses of the current state-of-the-art.

Automatic skin cancer detection and classification have been successfully developed in recent years with the method of artificial intelligence and deep learning. In [13], a fine-tuned vision transformer scheme is presented for skin tumor classification by using enhanced preprocessing, watershed-based segmentation, and hybrid feature extraction. The model was tested on the ISIC 2019 dataset with a performance of 99.81% accuracy, 96.65% precision, 98.21% sensitivity, and 97.42% F-measure, revealing the efficiency of transformer-based architectures for dermoscopic image analysis. A multi-class skin cancer identification system based on a deep network was proposed in [14], in which better Canny edge detection, along with optimized CNN and metaheuristic optimization, were employed to improve both lesion boundary detection and classification. Tested on the ISIC datasets, the proposed model can achieve approximately 99% accuracy, suggesting its promising power in multi-class skin lesions classification. In addition, an ensemble of deep learning models, combining SqueezeNet and InceptionResNetV2, was constructed with an enhanced Whale optimization algorithm in [15]. They achieved feature selection and classification robustness, obtaining 95.48% and 98.59% of accuracy, PH2 and Med-Node datasets, respectively.

Transfer learning-based techniques in deep learning have also been extensively investigated [16]. Modified a VGG16 model with more dense layers and some data augmentation for binary classification of skin cancer and reported an accuracy of 89.09% on a Kaggle dataset. Another work [17] used EfficientNet with transfer learning and Ranger optimizer for class balancing to enhance lesion classification and obtained an AUC score of 0.9681 on the ISIC datasets. These findings show the applicability of scalable CNNs for automated skin lesion characterization. Moreover, more sophisticated neural architectures have similarly been proposed to enhance diagnostic accuracy and interpretability. A mixed-order relation-aware RNN model trained by the Black-Winged Kite Algorithm was introduced in [18] for multi-class skin cancer detection. It was found that the method achieved superior performance with the accuracy of 99.89% and F1-score of 99.85%, respectively, resulting in more precise feature representation and classification precision. Furthermore, the cloud computing and AI-integrated healthcare systems were investigated in [19], where a cloud-based intelligent medical detection system with fuzzy neural network and whale optimization was suggested, which accentuates the significance of AI-assisted health monitoring and disease detection.

Optimization-oriented deep learning models have demonstrated great capabilities for enhancing the efficiency and accuracy of classification. Proposed an optimized CNN architecture based on particle swarm and bat algorithms for classifying skin lesions in the cloud-based diagnostic system to enhance the feature extraction and classification accuracy [20]. Traditional machine learning methods for cancer image classification have been reviewed in [21], which emphasized the disadvantage of former methods in contrast to deep learning-based approaches. Moreover, hybrid deep learning and machine learning methods have also been examined for enhanced diagnostic accuracy. In [22], an improved VGG19 network-based deep learning model integrated with the classical classifiers achieved higher classification accuracy and robustness in diagnosing skin cancer. Comparatively, experiments in [23] with machine learning and deep learning schemes using pre-trained CNN models and optimization techniques showed accuracy higher than 99% on ISIC datasets. These results validate the capability of deep learning models for dermoscopic image classification.

In [24], the authors addressed ensembling deep learning models with optimization techniques in a multi-class skin cancer classification framework through ensemble learning and White Shark Optimization, which achieved better diagnostic performance based on advanced pre-processing, feature extraction, and hyper-parameter tuning. The results showed improved classification performance and were insensitive to the parameter setting compared with other methods.

In recent studies, a format of pervasive healthcare monitoring and disease analysis has been approached with other emerging technologies. A recent critical review curation was published in [25], that addressed noninvasive wearable biosensors for diabetes monitoring, where the attention of the skin-interfaced sensors through continuous assessment of chemical and physiologic biomarkers such as glucose, cortisol, and heart rate were emphasized. Highlights of the study include multimodal sensor fusion, AI-based predictive analytics and closed-loop therapeutics for personalized and proactive care. It also covered instrumentation challenges such as sensor stability, data protection and regulation, and it concluded that next generation wearable biosensors would bring great benefit in enabling proactive diabetes management and enhancing patient outcomes.

In [26], an extensive review centred on lumpy skin disease (LSD), and emerging viral infection of cattle and buffaloes with considerable economic losses was reported. The research looked at the worldwide distribution, transmission dynamics, and molecular virology of the disease, finding widespread outbreaks throughout Africa, Asia, and across Europe. Sophisticated diagnostic technologies, including PCR, LAMP, and next-generation sequencing, were considered, as well as control measures like vaccination and vector elimination. The review identified critical hurdles such as a lack of diagnostic infrastructure, gaps in vector surveillance, and requirements for vaccine strategies to bolster integrated disease control efforts aimed at enhancing the health of livestock, ensuring food security, and economic stability.

The literature suggests that deep learning approaches of CAD systems have led to significant advances for skin cancer detection, but issues concerning model efficiency, scalability, generalization, and application into practice remain. Solving these problems would need the construction of a strong and effective deep learning-based system with dependable multi-class classification capacity, being computationally efficient, and feasible to be used in real-world applications. The findings of reviewed studies form a strong basis for the development of next-generation automated skin disease detection systems and to propel further research towards accurate, efficient, and clinically valuable diagnostic solutions.

### 3. METHOD

This section explains the general methodology applied to develop the proposed automated skin disease detection system. The methodology concerns the construction of a deep learning system that will accurately identify multiple classes of dermoscopic images conveying skin lesions. It contains dataset generation, image processing, model establishment, training, and testing. The goal is to develop a robust and highly accurate system to detect different kinds of skin lesions with good diagnostic performance.

Table 1. Description of Skin Lesion Types in the Study

Skin lesion Types	Description
Actinic Keratoses	Actinic keratoses are precancerous skin lesions. They are caused by long-term exposure to ultraviolet radiation.
Basal Cell Carcinoma	Basal cell carcinoma is the most common type of skin cancer. It is slow-growing and rarely spreads to other parts of the body.
Benign Keratosis-Like Lesions	Benign keratosis-like lesions are non-cancerous skin growths. They resemble actinic keratoses but are not precancerous.
Dermatofibroma	Dermatofibroma is a benign skin tumor. It is usually firm and raised and can be various colors.
Melanoma	Melanoma is a serious type of skin cancer. It can spread quickly to other parts of the body if not treated early.

The various categories of skin lesions analyzed in this study, with accompanying clinical descriptions, are shown in Table 1. The table consists of benign and malignant skin lesions, to enable multi-class classification. These are the lesion categories of interest and represent common skin conditions included in training and testing the model. Insight into the characteristics of individual categories of lesions is useful in better recognizing and automatically detecting skin diseases.

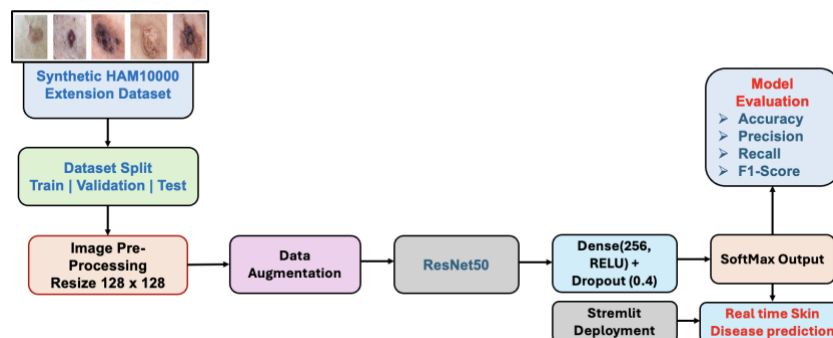


Figure 1. Framework of the introduced deep learning-based skin disease detection system

The whole process of the proposed skin disease diagnosis system is presented in Figure 1. The process starts with the synthetic extension dataset Synthetic HAM10000 distributed over training, validation, and test sets in order to evaluate models. The input images are pre-processed, scaled to 128×128 pixels, and augmented to prevent over-fitting and improve the model's generalization. The processed images are fed to the ResNet50 deep network along with a fully connected and dropout layer for feature extraction and classification. Lastly, the model's predictions are made with a SoftMax output layer, and its performance results are based on accuracy, precision, recall, and F1 score, while deployed through Streamlit for real-time skin disease prediction.

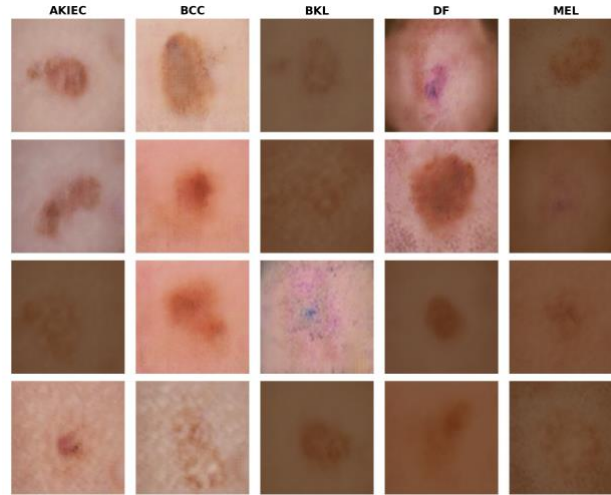


Figure 2. Samples of Dermoscopic Images from Different Skin Disease Categories

Figure 2 shows some dermoscopic image examples of the five types of skin diseases studied here: Actinic Keratoses (AKIEC), Basal Cell Carcinoma (BCC), Benign Keratosis-like Lesions (BKL), Dermatofibroma (DF), and Melanoma (MEL). Actinic Keratoses (AKIEC) are lesions of the skin, precancerous to most common cancers on the skin, and are predominantly associated with exposure to ultraviolet radiation, which can lead to cancer if left untreated. Basal Cell Carcinoma BCC is the most frequent skin tumour; generally being low-growing and with little metastatic potential, but, nevertheless, early stages of diagnosis are needed for appropriate treatment. Benign Keratosis-like (BKL) are non-cancerous growths, which mimic precancerous lesions and rarely pose a serious threat to health. Dermatofibroma (DF) is a benign skin tumour visible as a small, hard bump and usually not harmful. Melanoma (MEL), on the other hand, is the worst form of skin cancer and can easily spread to other organs if it's not caught in time. These representative images of various lesion types demonstrate the appearance differences and are employed to train and test the deep learning-based skin disease classification method.

### 3.1. Dataset Preparation and Input Representation

Let us represent the training data as shown in equation (1).

$$D = \{(x_i, y_i)\}_{i=1}^N \quad (1)$$

where N is the number of total training samples. In the dataset, each item comprises an input image  $x_i$  and a label  $y_i$ . Here,  $x_i \in \mathbb{R}^{128 \times 128 \times 3}$  denotes a dermoscopic skin lesion image rescaled with spatial size of 128×128 pixels and three-dimensional color channels (RGB). The corresponding label  $y_i \in \mathbb{R}^C$ . The total number of skin disease classes is denoted as C. In such encoding, the class label vector has a value of 1 at only one index position and 0 elsewhere. This equation represents a mathematical form for supervised learning, in which the model aims to find an optimal mapping from the input image space to its corresponding class label space for the correct discrimination of skin diseases.

### 3.2. Feature Extraction Using ResNet50

In the proposed framework, a pretrained ResNet50 network is utilized as a deep feature extractor to extract discriminative features from input dermoscopic images. The feature extraction can be written as Equation (2).

$$F = f_{\theta}(x) \quad (2)$$

where  $x$  is the input image,  $f_{\theta}(\cdot)$  is the nonlinear mapping introduced by the ResNet50 convolutional layers with weights  $\theta$ , and  $F = \in \mathbb{R}^{H \times W \times D}$  are feature maps obtained. Here, H and W are the spatial dimensions of the feature map, and D stands for depth, as in the number of channels that encode various learned visual patterns like texture, color, or lesion structure. In this article, we use a pretrained model, ResNet50, for transfer learning and freeze its convolutional base. Hence, the training of the pretrained parameters that are learned from a complex dataset is frozen. This enables the network to make use of generic visual features learned earlier, which lowers the computational burden and avoids overfitting on the skin lesion dataset.

### 3.3. Global Average Pooling

The high-dimensional feature maps output by the convolutional network, following feature extraction, are compressed into a compact feature vector by applying the Global Average Pooling (GAP) operation. From a mathematical point of view, GAP takes the average value across spatial dimensions of each feature channel shown in equation (3).

$$z_k = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F_{i,j,k} \quad (3)$$

where  $F_{i,j,k}$  is the activation at spatial location  $(i, j)$  in the  $k$ -th feature channel and H and W are height and width of the feature map respectively. The resulting vector  $z = \in \mathbb{R}^D$  consists of one value per channel, summarizing the global appearance of learned visual patterns such as lesion texture, color distribution, and structural cues. GAP helps reduce the number of parameters as compared to fully connected flattening, prevents overfitting, and retains the most vital global information from the feature maps. All these properties make it particularly appropriate for tasks such as image classification.

### 3.4. Fully Connected Layer

The pooled feature vector from the Global Average Pooling is fed into a fully connected (dense) layer, where we first apply a linear transformation and then an activation function. This operation can be written as equation (4).

$$h = \sigma(W_1 z + b_1) \quad (4)$$

where  $z$  is the pooled feature vector,  $W_1$  is a learnable weight matrix,  $b_1$  is a bias vector and  $\sigma(\cdot)$  denotes activation function. In this architecture, we use the rectified linear unit (ReLU) activation as follows in equation (5).

$$ReLU(x) = \max(0, x) \quad (5)$$

The input and output of the fully connected layer are features to learn higher-level representations through combination, and the ReLU activation also adds nonlinearity to the network. This allows the model to learn more complex and discriminative patterns from various types of skin lesions, enhancing its ability to differentiate between classes in classification.

### 3.5. Dropout Regularization

To prevent the model from overfitting and to enhance its generalization capability, a dropout layer is used following the fully connected layer. Dropout does this by turning off a random selection of neurons while training, which prevents the network from becoming overly reliant on particular features. In mathematical terms, this reads in equation (6).

$$\hat{h} = h \cdot m, m \sim Bernoulli(p) \quad (6)$$

where  $h$  is the input feature vector from the previous layer,  $m$  is a binary mask with probability  $p$  ( $m$ ) that follow Bernoulli distribution. The dropout parameter value in this study is 0.4, which implies that each neuron is used with the probability  $p = 0.6$  to be activated and the probability of deactivation during the training process. The model is less prone to co-adaptation of features and learns more robust and generalized representations, which in turn lead to better classification of unseen skin-lesion images due to the random dropout.

### 3.6. SoftMax Classification

The output layer of the network computes class probabilities via SoftMax on the learned feature representation. In mathematical terms this operation is represented by equation (7).

$$\hat{y}_c = \frac{e^{W_{2,c} \hat{h} + b_{2,c}}}{\sum_{k=1}^C e^{W_{2,k} \hat{h} + b_{2,k}}} \quad (7)$$

where  $\hat{h}$  is the feature vector after dropout,  $W_2$ ,  $b_2$  are the weight and bias of the final fully-connected layer, respectively,  $C$  is the total number of skin disease classes. It is the predicted probability that a given input image belongs to class (c). The output scores are normalized into a probability distribution by the SoftMax function, where probabilities for all classes add up to one. This lets the model label each input image with the most likely class and, therefore, can be applied to a multiclass skin disease classification problem.

### 3.7. Loss Function (Categorical Cross-Entropy)

The categorical cross-entropy loss function is used to train the network, which measures how well it is doing of predicting class probabilities in comparison with true class labels. This loss is defined as in equation (8).

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

where  $C$  is the number of classes,  $y_c$  is the true class label one-hot encoded for class (c), and  $\hat{y}_c$  is the predicted probability for each from the Softmax layer. As the ground truth label vector consists of a single 1 value for the correct class and 0 everywhere else, its loss reflects how well the predicted probability for the correct class corresponds to this true label. Categorical cross-entropy bonus the erroneous predictions that have a low probability when they are assumed to be the ground truth class in order to push the network's parameters to update in such a way that it increases the probability of other predicted classes for the correct class. Training can be performed using gradient descent to minimize this loss function and encourage the model to predict accurate and stable skin lesion classification.

### 3.8. Optimization Using Adam

The model parameters are learned with the Adam optimizer that updates network weights step by step to minimize the loss. The update rule for the parameter is described as in equation (9).

$$\theta_{t+1} = \theta_t - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \quad (9)$$

where  $\theta_t$  and  $\theta_{t+1}$  are the model parameters at the current iteration and at the next iteration, respectively, and  $\alpha$  is the learning rate 0.0005 in our experiments  $\hat{m}_t$  is the bias-corrected first moment estimate (mean of gradients),  $\hat{v}_t$  is the bias-corrected second-moment estimate (variance of gradients), and  $\epsilon$  to avoid division by zero.

Adam has the benefits of both momentum and adaptive learning rate methods; the first moment (mean) estimate is utilized to smooth updates of the gradient, and the second moment (variance) estimate can adaptively adjust the learning rate for each weight. This leads to quick convergence, smooth training, and effective optimization, which makes Adam an excellent choice for deep learning-based skin lesion classification.

### 3.9. Performance Metrics

The proposed skin disease classification model is evaluated using the conventional performance measures, such as Accuracy, Precision, Recall, and F1-score from the confusion matrix. These criteria give an overall measure of the prediction ability of the model.

Accuracy is the global ratio of correctly predicted samples and is described by equation (10).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

where TP, TN, FP, and FN are true positives, true negatives, false positives, and false negatives, respectively. It is a measure of how many of the actual positives our model captures through labeling it as positive, and ranges at a certain scale, compared to precision.

Precision is a measure of the proportion of true positive cases, and is defined as follows in equation (11). It indicates the percentage of instances that the samples were predicted as a certain skin disease, which is correct, referring to the success rate and representing the dependability of positive classification.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (11)$$

The recall (or sensitivity) is the measurement of how well a model returns true positive cases, and is defined as in equation (12). Higher recall means that the model detects most of the actual disease cases.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (12)$$

F1-score is the harmonic mean of precision and recall, and it is useful for providing a single measure when both precision and recall are equally important, as shown in equation (13).

$$F1 - score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (13)$$

It measures the balance between precision and recall to merge into a single overall score, so that the model will not only do well in one class but poorly with another. These measurements provide a comprehensive analysis perspective of the classification ability and consistency of the proposed deep learning-based skin disease detection system.

#### 4. RESULTS AND DISCUSSION

In this section, we evaluate the efficacy of our proposed model via experimental results and performance analysis. The dataset was further split into 5,488 training images, 1,644 validation images, and an independent testing set to provide accurate and unbiased evaluation. Performance of the tuned model was evaluated with standard metrics to investigate its performance for different skin lesion types. The results obtained show the effectiveness of the proposed method for precise detection and classification of different skin diseases.

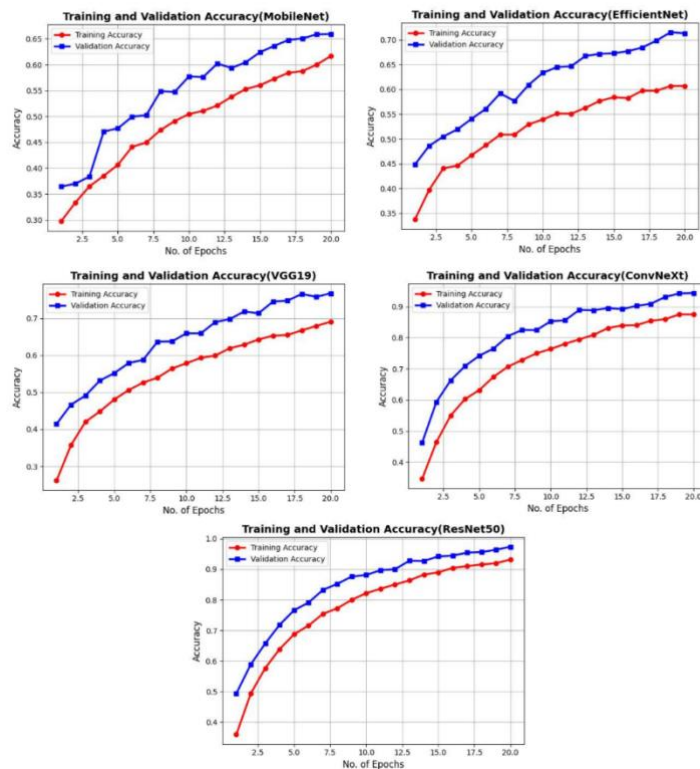


Figure 3. Comparison of Training and Validation Accuracy between Different Deep Learning Models

Figure 3 shows the prediction accuracy curves in the training and validation datasets using five selected DNNs -MobileNet, EfficientNet, VGG19, ConvNeXt and ResNet50- for the different learning epochs. The curves indicate an increase in accuracy for both training and validation data with the number of epochs, which suggests reaching convergence and effective learning. First among the architectures compared is ResNet50, which performs best with fastest convergence and smallest gap between training and validation accuracy, indicating better generalization and lesser over-fitting. MobileNet and EfficientNet perform relatively lower, and VGG19 and ConvNeXt display some improvement, but still not better than ResNet50. An observation illustrates the effectiveness of using ResNet50 in facilitating accurate skin disease classification.

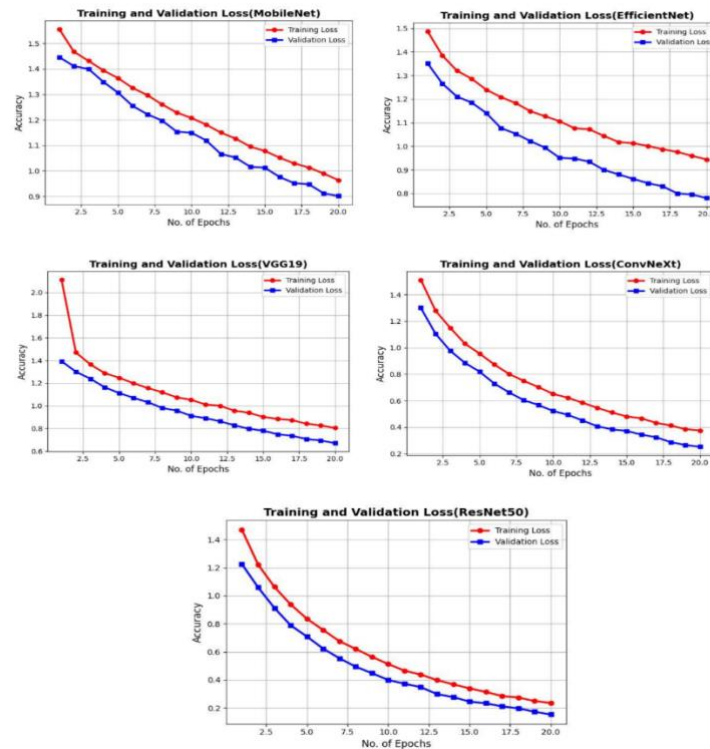


Figure 4. Training and Validation Loss Comparison for Various Deep Learning Models

The training and validation loss curve of five deep learning architectures, MobileNet, EfficientNet, VGG19, ConvNeXt, and ResNet50, with various training epochs, is illustrated in Figure 4. All curves decrease monotonically with the number of epochs of training and validation progress, which reflects good learning and optimization for these models. Among all architectures, ResNet50 presents the fastest convergence in terms of loss reduction and also obtains the lowest final loss value, which indicates better model convergence and generalization. ConvNeXt and VGG19 have steady but higher loss decreases than ResNet50, whereas MobileNet and EfficientNet are relatively slow in performance improvement. The close distance between training and validation loss in ResNet50 means less overfitting and more stable learning behavior, which reflects the best performance of ResNet50 as a skin disease classifier on the proposed method.

Figure 5 shows the confusion matrices of multi-class skin disease classification using five deep learning models: MobileNet, EfficientNet, VGG19, ConvNext, and ResNet50. Each confusion matrix compares the distribution of true class labels versus predicted labels among the five classes: Actinic Keratoses (AKIEC), Basal Cell Carcinoma (BCC), Benign Keratosis-like Lesions (BKL), Dermatofibroma (DF), and Melanoma (MEL). The on-diagonal elements correspond to the frequency of true classifications, while off-diagonal elements represent samples that have been misclassified. Looking at all models, ResNet50 gives the most correct predictions and the least misclassification, showing better classification performance and robustness. ConvNeXt also achieves high performance but is slightly inferior to ResNet50, while VGG19, EfficientNet, and MobileNet present relatively higher confusion among visually similar lesion classes such as BKL and MEL. Overall, the figure demonstrates that ResNet50 performed well in accurately discriminating between various skin disease categories with better consistency and less errors.

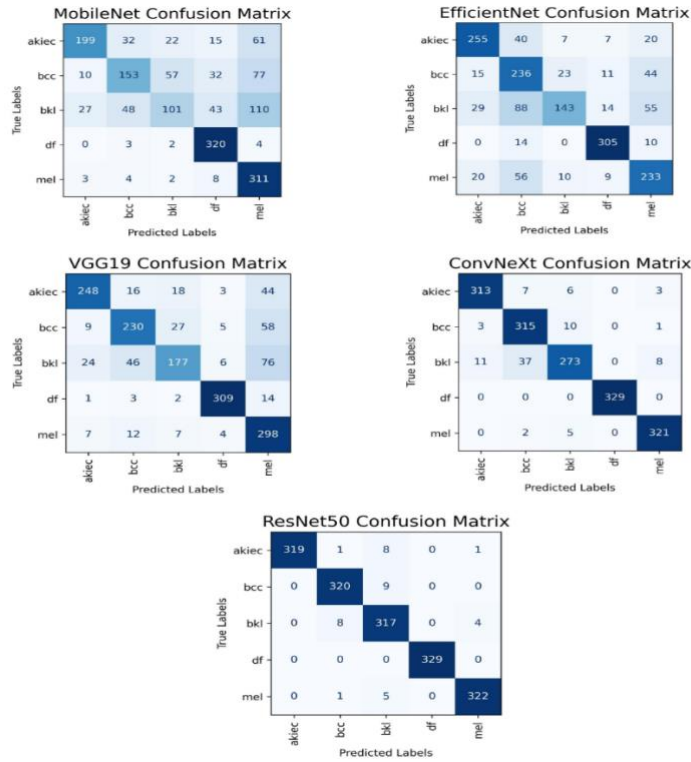


Figure 5. Confusion Matrix Comparison of Skin Disease Classification Deep Learning Models

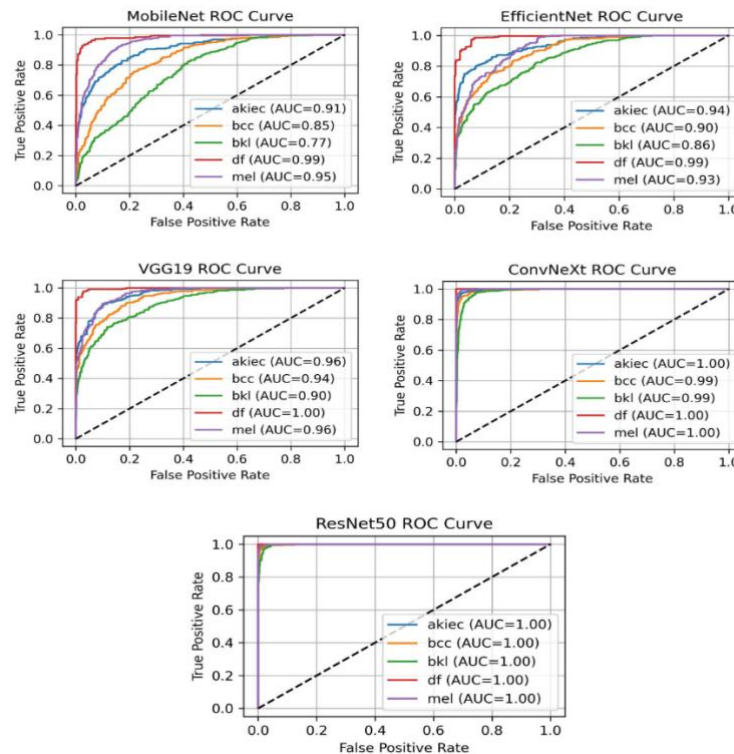


Figure 6. ROC Curves and AUC Scores for Comparative Models

The ROC curves and AUC values for each of the models, including MobileNet, EfficientNetB0, VGG19, ConvNeXt, and the proposed ResNet50 model for all lesion classes, are shown in Figure 6. The comparison models perform with different levels of distinguishability; MobileNet and EfficientNet obtain moderate discrimination, and AUC is relatively low, while VGG19 has better separability capability. ConvNeXt achieves impressive AUC values that are close to perfect over most classes. For all the lesion categories, the proposed ResNet50 model obtains an AUC score of 1.00, which means that we are able to

completely separate positives from negatives. This great performance could be evidence that ResNet50 feature extraction and decision boundary learning can outperform those other models.

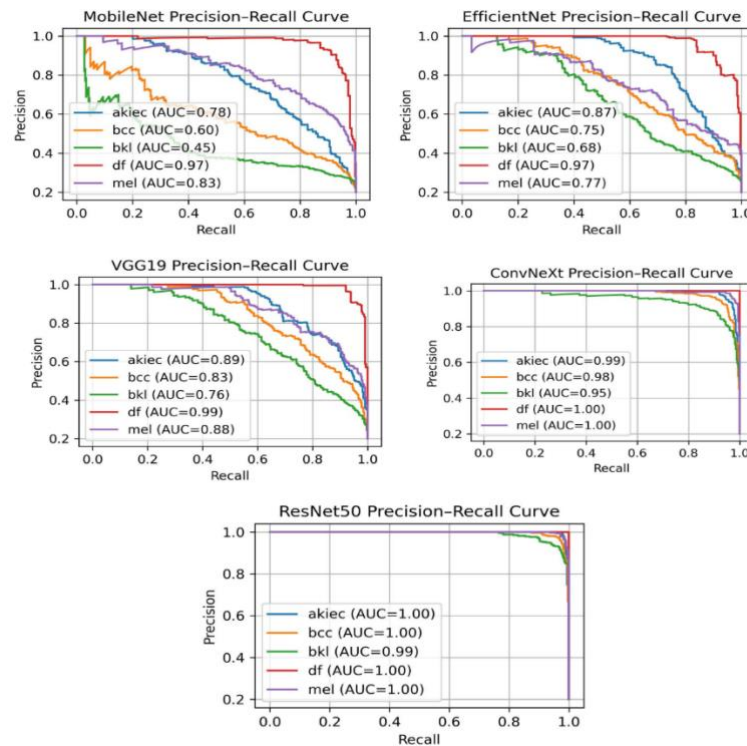


Figure 7. Precision-Recall Curves and AUC Scores for Comparative Models

The precision-recall curves and their AUC measurements for MobileNet, EfficientNetB0, VGG19, ConvNeXt, and the proposed ResNet50 model across all lesion types are shown in Figure 7. Comparison architectures exhibit diverse precision-recall behaviour, with MobileNet and EfficientNet getting moderate performance and VGG19 obtaining stronger but unstable class separation. ConvNeXt substantially outperforms this, achieving high AUC across most of the classes. ResNet50 model achieves very good or perfect AUCs for all the lesions, suggesting extreme discrimination of positive samples and negative samples even when the classes are imbalanced. The ResNet50 curves are well-separated, therefore implying a satisfactory precision at high recall levels and reinforcing the idea that this architecture is more apt at finding true positives without generating false alarms compared to the other models.

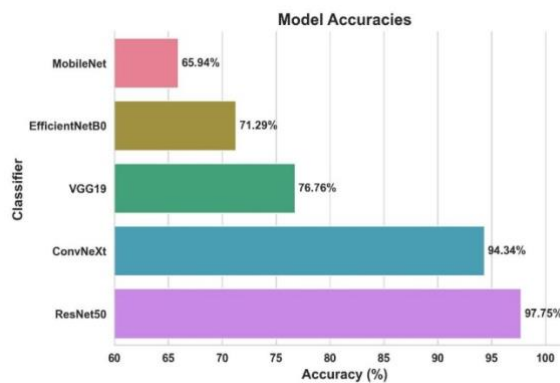


Figure 8. Accuracy Comparison of Deep Learning Models

Figure 8 shows a comparison of classification accuracy performance between what can be afforded by five deep learning architectures: MobileNet, EfficientNetB0, VGG19, ConvNeXt and ResNet50. From the enclosed bar chart, we can see that the use of MobileNet resulted in an accuracy of 65.94%, followed by EfficientNetB0 with 71.29%, and VGG19 at 76.76%, suggesting a fair performance in classification results.

ConvNeXt performed impressively well with an accuracy of 94.34%, indicating its very good feature learning ability. The Highest accuracy 97.75% was achieved by ResNet50, which surpassed by a large margin the other network architecture. The superiority of the ResNet50 model is shown and demonstrates its advantages in accuracy, robustness, and generalization on multi-class skin disease classification.

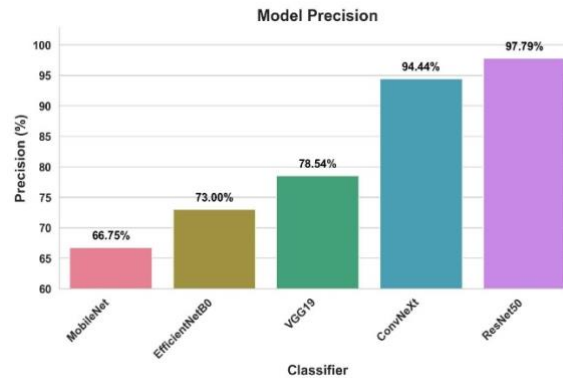


Figure 9. Precision Comparison of Deep Learning Models

Figure 9 shows a comparison of precision by five deep learning models—MobileNet, EfficientNetB0, VGG19, ConvNeXt, and ResNet50 for multi-class skin disease classification. Precision is the ratio of the number of true positive cases divided by all predicted positives and reflects how trustworthy model predictions are. MobileNet obtained the precision of 66.75%, EfficientNetB0 with prediction performance, and VGG19 with 78.54% moderate classification reliability. This is due to ConvNeXt, which has a precision of 94.44%, showing that it could learn a very discriminant feature. ResNet50 obtained the best precision of 97.79% in comparison to all models, showing better performance with a low number of false positive predictions. These findings validate that the ResNet50 model is the most robust and precise with respect to classification compared to other architectures examined.

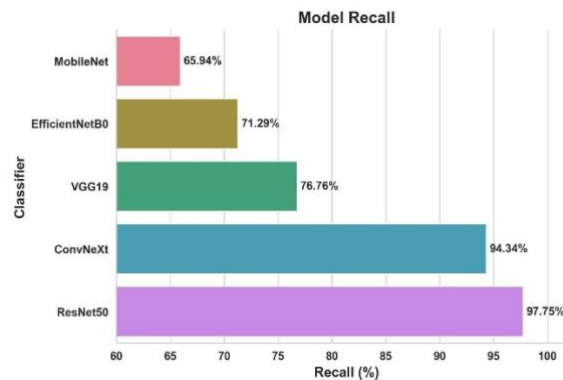


Figure 10. Recall Comparison of Deep Learning Models

Figure 10 compares the recall values of MobileNet, EfficientNetB0, VGG19, ConvNeXt, and ResNet50 deep learning architectures in multi-class skin disease classification. Recall determines the model's ability to identify all relevant actual positives accurately, which indicates its performance in assessing the true prevalence of a disease. MobileNet had the least recall, 65.94%, while EfficientNetB0 and VGG19 had 71.29% and 76.76%, respectively, indicating fair detection performance. ConvNeXt has an average recall of 94.34%, which means the model is good at identifying true positives. ResNet50 had the highest recall of 97.75%, indicating high precision and a small number of false negatives. These findings demonstrate that ResNet50 can detect and categorize skin diseases more accurately than other tested models.

Figure 11 compares F1-scores between our five deep learning models: MobileNet, EfficientNetB0, VGG19, ConvNeXt, and ResNet50 for multi-class skin disease classification. The F1-score is the harmonic mean of precision and recall, so it's a more conservative estimate of model performance, particularly if false/false negatives are equally important. MobileNet reached an F1-score of 63.74%, EfficientNetB0 with 70.87%, and VGG19 at 76.58%, suggesting that a moderate classification performance was accomplished. ConvNeXt presented a remarkable performance with an F1-score of 94.30%, indicating good overall predictive power. Taking all the models into consideration, ResNet50 obtained the highest F1-score 97.76%,

indicating better balance between precision and recall, and validating it as the best model for accurate multi-class skin disease classification.

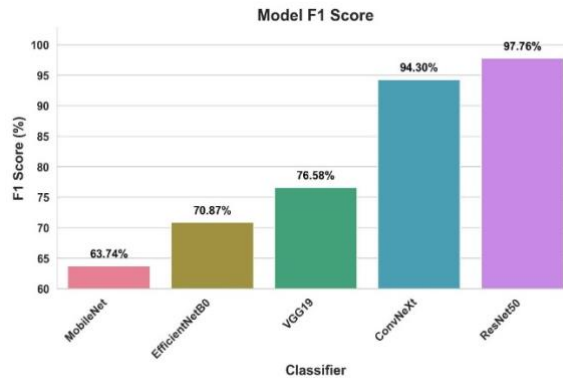


Figure 11. F1-Score Comparison of Deep Learning Models

Table 2. Class-wise Evaluation of ResNet50 for the Prediction of Skin Disease

Class	Precision	Recall	F1-score	Support
Actinic Keratoses (AKIEC)	1.00	0.97	0.98	329
Basal Cell Carcinoma (BCC)	0.97	0.97	0.97	329
Benign Keratosis-like Lesions (BKL)	0.94	0.96	0.95	329
Dermatofibroma (DF)	1.00	1.00	1.00	329
Melanoma (MEL)	0.98	0.98	0.98	328

The class-wise classification performance of the proposed ResNet50 model on skin disease dataset using Precision, Recall and F1-score is reported in Table 2. The model performs well across all five skin lesion categories. Dermatofibroma (DF) was perfectly classified with 100% precision, recall, and F1-score; and no instance of misclassification resulted for this class. AKIEC and MEL also performed very well receiving F1-score on 0.98 accurate spectral detection and discrimination. The BCC model set the point of perfect balance of precision and recall apart from 0.97, demonstrating credible predictions. Benign Keratosis-like Lesions (BKL) reached lower performance with an F1-score of 0.95 caused by a small confusion with visually related lesions. These results validate that the ResNet50 model predicts skin disease categories with high accuracy and consistency.

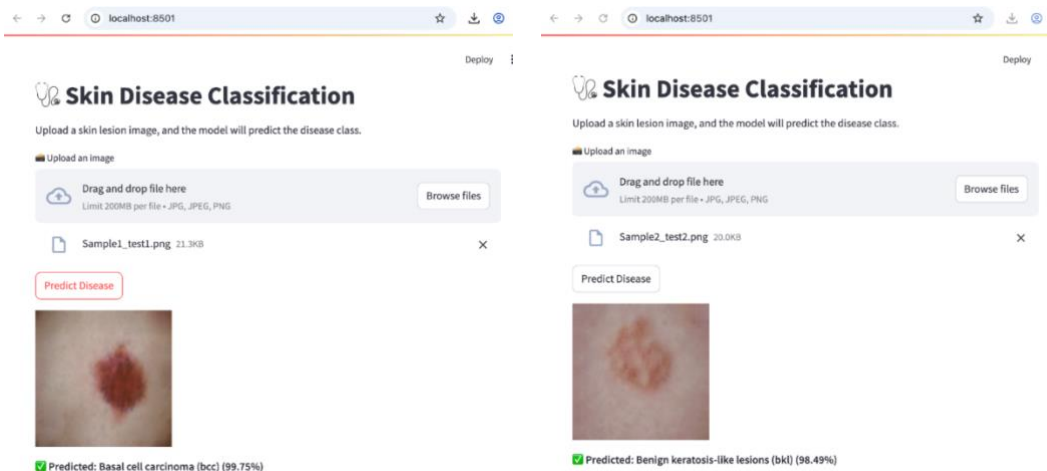


Figure 12. Streamlit based Real-time skin disease prediction using ResNet50

Figure 12 presents a real-time application and testing of the proposed skin disease classification model based on ResNet50 using a Streamlit web application. Dematic WebApp has a user interface for users

to upload dermoscopic skin lesion images, where the pre-trained model automatically predicts disease class and confidence score. In the first scenario, the input image is classified as Basal Cell Carcinoma (BCC) with 99.75% prediction confidence. In the second one, we can see everything is working as expected, where the systems correctly diagnose it as Benign Keratosis-like Lesions (BKL) and 98.49% confident about that. These outcomes demonstrate the efficiency, stability, and applicability of the proposed model in real-time skin disease detection. The deployed Streamlit view indicates that the trained ResNet50 model could be implemented into an interactive clinical decision-support system, which would provide rapid and reliable classification of skin lesions for automated diagnosis support.

## 5. CONCLUSION

In this work, we have proposed an efficient deep learning-based framework to automatically classify skin diseases from dermoscopic images. Several state-of-the-art CNN models, such as MobileNet, EfficientNetB0, VGG19, ConvNeXt, and ResNet50 were developed and evaluated in order to determine the best model capable of accurately classifying skin lesions into multiple categories. Experimental results showed that the extracted deep feature, in combination with transfer learning, achieves great improvement on classification performance and stability. Since an unbalanced distribution of samples may mislead the performance evaluation, we use accuracy, which is an inherent property of weights, and balance our model when calculating average classification accuracy among skin disease classes. Among all tested models, ResNet50 also gives the best overall performance, with an average classification accuracy of 97.75%; concurrently with high precision, recall, and F1-score for each skin disease. This observation is strengthened by the fact that, as per the confusion matrix and class-wise results, our proposed ResNet50-based model discriminates effectively against visually close skin lesions with a low level of misclassification. In addition, and for its practical applicability as an intelligent clinical decision-support tool, the model was put into operation with a Streamlit web-app platform for real-time prediction of skin disease. The proposed method offers a robust, accurate, and efficient solution to automatic skin disease detection and classification that can aid dermatologists in the early diagnosis process and improve healthcare quality. In future work, the model can be augmented by feeding larger and more diverse datasets into more sophisticated attention-based architectures and explainable AI methods for increasing diagnostic accuracy, interpretability, and clinical practicality.

## DATA AVAILABILITY STATEMENT

The results are generated using the dataset “Synthetic HAM10000 Extension Dataset with Skin Tone Balance” available in “IEEE Dataport” with the doi: 10.21227/51cx-yy74

## CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest to this work.

## REFERENCES

- [1] P. Hermosilla, R. Soto, E. Vega, C. Suazo, and J. Ponce, “Skin Cancer Detection and Classification Using Neural Network Algorithms: A Systematic Review,” *Diagnostics*, vol. 14, no. 4, p. 454, Feb. 2024, <https://doi.org/10.3390/diagnostics14040454>.
- [2] K. Anup Kumar and C. Vanmathi, “Optimization driven model and segmentation network for skin cancer detection,” *Computers and Electrical Engineering*, vol. 103, p. 108359, Oct. 2022, <https://doi.org/10.1016/j.compeleceng.2022.108359>.
- [3] A. K. K., S. T.Y., S. T. Ahmed, S. K. Mathivanan, S. Varadhan, and M. A. Shah, “Trained neural networking framework based skin cancer diagnosis and categorization using grey wolf optimization,” *Scientific Reports*, vol. 14, no. 1, p. 9388, Apr. 2024, <https://doi.org/10.1038/s41598-024-59979-4>.
- [4] T. Imran, A. S. Alghamdi, and M. S. Alkathiri, “Enhanced Skin Cancer Classification using Deep Learning and Nature-based Feature Optimization,” *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12702–12710, Feb. 2024, <https://doi.org/10.48084/etasr.6604>.
- [5] J. Jaculin Femil and T. Jaya, “An Efficient Hybrid Optimization for Skin Cancer Detection Using PNN Classifier,” *Computer Systems Science and Engineering*, vol. 45, no. 3, pp. 2919–2934, 2023, <https://doi.org/10.32604/csse.2023.032935>.
- [6] N. Zhang, Y.-X. Cai, Y.-Y. Wang, Y.-T. Tian, X.-L. Wang, and B. Badami, “Skin cancer diagnosis based on optimized convolutional neural network,” *Artificial Intelligence in Medicine*, vol. 102, p. 101756, Jan. 2020, <https://doi.org/10.1016/j.artmed.2019.101756>.
- [7] D. Adla, G. V. R. Reddy, P. Nayak, and G. Karuna, “Deep learning-based computer aided diagnosis model for skin cancer detection and classification,” *Distributed and Parallel Databases*, vol. 40, no. 4, pp. 717–736, Dec. 2022, <https://doi.org/10.1007/s10619-021-07360-z>.
- [8] K. Ali, Z. A. Shaikh, A. A. Khan, and A. A. Laghari, “Multiclass skin cancer classification using EfficientNets – a first step towards preventing skin cancer,” *Neuroscience Informatics*, vol. 2, no. 4, p. 100034, Dec. 2022, <https://doi.org/10.1016/j.neuri.2021.100034>.

- [9] Priyanka Tyagi and S.K. Manju bargavi, "Using Federated Artificial Intelligence System of Intrusion Detection for IoT Healthcare System Based on Blockchain," *International Journal of Data Informatics and Intelligent Computing*, vol. 2, no. 1, pp. 1–10, Mar. 2023, <https://doi.org/10.59461/ijdiic.v2i1.42>.
- [10] M. M. Musthafa, M. T R, V. K. V, and S. Guluwadi, "Enhanced skin cancer diagnosis using optimized CNN architecture and checkpoints for automated dermatological lesion classification," *BMC Medical Imaging*, vol. 24, no. 1, p. 201, Aug. 2024, <https://doi.org/10.1186/s12880-024-01356-8>.
- [11] L. Zhang, H. J. Gao, J. Zhang, and B. Badami, "Optimization of the Convolutional Neural Networks for Automatic Detection of Skin Cancer," *Open Medicine*, vol. 15, no. 1, pp. 27–37, Jan. 2020, <https://doi.org/10.1515/med-2020-0006>.
- [12] A. Raza, A. Ali, S. Ullah, Y. N. Anjum, and B. Rehman, "Optimizing skin cancer screening with convolutional neural networks in smart healthcare systems," *PLOS ONE*, vol. 20, no. 3, p. e0317181, Mar. 2025, <https://doi.org/10.1371/journal.pone.0317181>.
- [13] R. P. Desale and P. S. Patil, "An efficient multi-class classification of skin cancer using optimized vision transformer," *Medical & Biological Engineering & Computing*, vol. 62, no. 3, pp. 773–789, Mar. 2024, <https://doi.org/10.1007/s11517-023-02969-x>.
- [14] J. S. Thanga Purni and R. Vedhapriyavadhana, "EOSA-Net: A deep learning framework for enhanced multi-class skin cancer classification using optimized convolutional neural networks," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 3, p. 102007, Mar. 2024, <https://doi.org/10.1016/j.jksuci.2024.102007>.
- [15] A. Majid, M. A. Alrasheedi, A. A. Alharbi, J. Allohibi, and S.-W. Lee, "Modified Whale Optimization Algorithm for Multiclass Skin Cancer Classification," *Mathematics*, vol. 13, no. 6, p. 929, Mar. 2025, <https://doi.org/10.3390/math13060929>.
- [16] V. Anand, S. Gupta, A. Altameem, S. R. Nayak, R. C. Poonia, and A. K. J. Saudagar, "An Enhanced Transfer Learning Based Classification for Diagnosis of Skin Cancer," *Diagnostics*, vol. 12, no. 7, p. 1628, Jul. 2022, <https://doi.org/10.3390/diagnostics12071628>.
- [17] J. S M, M. P, C. Aravindan, and R. Appavu, "Classification of skin cancer from dermoscopic images using deep neural network architectures," *Multimedia Tools and Applications*, vol. 82, no. 10, pp. 15763–15778, Apr. 2023, <https://doi.org/10.1007/s11042-022-13847-3>.
- [18] G. S. Uthayakumar, M. Yaramadhi, T. Marimuthu, and T. R. V. Lakshmi, "Optimized Mixed-Order Relation-Aware Recurrent Neural Networks based CAD Model for Skin Cancer Detection and Classification," *Knowledge-Based Systems*, vol. 315, p. 113222, Apr. 2025, <https://doi.org/10.1016/j.knosys.2025.113222>.
- [19] Arvind Kumar Shukla and V. Suresh Kumar, "Cloud Computing with Artificial Intelligence Techniques for Effective Disease Detection," *International Journal of Data Informatics and Intelligent Computing*, vol. 2, no. 1, pp. 32–41, Mar. 2023, <https://doi.org/10.59461/ijdiic.v2i1.45>.
- [20] M. R. Mundada *et al.*, "Skin Cancer Prediction by Incorporating Bio-inspired Optimization in Deep Neural Network," *SN Computer Science*, vol. 5, no. 8, p. 1127, Dec. 2024, <https://doi.org/10.1007/s42979-024-03501-0>.
- [21] Ashish Kumar Pandey and Prabhdeep Singh, "A Systematic Survey of Classification Algorithms for Cancer Detection," *International Journal of Data Informatics and Intelligent Computing*, vol. 1, no. 2, pp. 34–50, Dec. 2022, <https://doi.org/10.59461/ijdiic.v1i2.32>.
- [22] I. A. Kandhro *et al.*, "Performance evaluation of E-VGG19 model: Enhancing real-time skin cancer detection and classification," *Heliyon*, vol. 10, no. 10, p. e31488, May 2024, <https://doi.org/10.1016/j.heliyon.2024.e31488>.
- [23] A. Magdy, H. Hussein, R. F. Abdel-Kader, and K. A. El Salam, "Performance Enhancement of Skin Cancer Classification Using Computer Vision," *IEEE Access*, vol. 11, pp. 72120–72133, 2023, <https://doi.org/10.1109/ACCESS.2023.3294974>.
- [24] R. V. Arumugam and S. Saravanan, "Automated multi-class skin cancer classification using white shark optimizer with ensemble learning classifier on dermoscopy images," *Multimedia Tools and Applications*, vol. 84, no. 8, pp. 4857–4879, Mar. 2024, <https://doi.org/10.1007/s11042-024-18973-8>.
- [25] A. Sedighi, T. Kou, H. Huang, and Y. Li, "Noninvasive On-Skin Biosensors for Monitoring Diabetes Mellitus," *Nano-Micro Letters*, vol. 18, no. 1, p. 16, Dec. 2026, <https://doi.org/10.1007/s40820-025-01843-9>.
- [26] P. Sharma and D. Kumar, "Global expansion of lumpy skin disease: transmission trends, economic consequences, and preventive measures," *Veterinary Research Communications*, vol. 50, no. 1, p. 22, Feb. 2026, <https://doi.org/10.1007/s11259-025-10954-y>.

## BIOGRAPHIES OF AUTHORS



**Sujatha Krishna** received the B.E. degree in Computer Science and Engineering from S.J.C. Institute of Technology, affiliated with Visvesvaraya Technological University (VTU), India, and the M.Tech. degree in Computer Science and Engineering from REVA Institute of Technology and Management, also affiliated with Visvesvaraya Technological University (VTU), India. She earned her Ph.D. degree in Computer Science and Engineering from REVA University, India. In recognition of her academic excellence, she was awarded the "Best Outgoing Student in Academics" during her M.Tech. program. She has consistently demonstrated a strong commitment to teaching, mentoring, and academic coordination throughout her career. She is currently serving as a Lecturer at the University of Technology and Applied Sciences – Shinas, Oman, and has over 15 years of teaching experience at both undergraduate and postgraduate levels. Her research

interests include big data analytics, data mining, data warehousing, machine learning, privacy-preserving algorithms, and secure data sharing in big healthcare data environments. She can be contacted at email: Sujatha.Krishna@utas.edu.om.



**Osamah Ibrahim Khalaf**, professor at Al-Nahrain University. I have a strong research background in computer science and information technology, 17 years of university-level teaching experience, over 141 ISI-indexed publications, and numerous international conference presentations. I hold two Australian patents, have completed over 1300 peer reviews, and have received multiple awards for my innovative work. I also have an h-index of 61. I am a highly accomplished researcher and academic with a strong international reputation due to my contributions to computer science. My work has been widely acknowledged through prestigious awards, patents, publications, and leadership roles in global academic and research communities, including committee and editorial roles for international conferences and journals. Throughout my career, I have contributed to critical advancements in embedded and Real-Time Communication Systems, wireless sensor networks, and healthcare security. My work on a water resource management system for a sustainable environment (400+ citations) has optimized urban water use through smart systems and green infrastructure. In energy-efficient clustering for wireless sensor networks (500+ citations), I developed algorithms to extend network lifespan and enhance performance for applications like environmental monitoring and smart agriculture. Additionally, my research on secure healthcare systems using machine learning (200+ citations) leverages AI-driven approaches, such as support vector machines, to protect medical data and improve patient trust. These accomplishments not only demonstrate the impact of my research across multiple disciplines but also underscore my ability to drive continued innovation in the World, contributing to advancements in sustainability, technology, and public health. In the coming years, I intend to extend my research to develop more efficient, interpretable, Embedded and Real-Time Communication Systems and ethical AI models, including federated learning, generative AI (e.g., large language models and diffusion models), AI for scientific discovery, quantum machine learning, error correction, and hybrid quantum-classical systems. A major challenge in the field is improving the robustness and fairness of AI systems, particularly in high-stakes decision-making areas such as healthcare, criminal justice, and hiring. My research aims to develop methods to audit and mitigate biases in datasets and models while maintaining performance, design interpretable AI architectures without sacrificing accuracy, and enhance adversarial robustness through novel training paradigms or formal verification. My work aligns with the growing demand for ethical, trustworthy AI in the World where regulatory frameworks and industry needs are rapidly evolving. This research is highly relevant to the World which invests heavily in quantum technologies, as breakthroughs have revolutionary implications for optimization, cryptography, and materials science. He can be contacted at email: usama81818@nahrainuniv.edu.iq.