**RESEARCH PAPER**

# Comparative Study of Deep Learning Models for Distinction on Multi-Class Skin Disease Image Classification

**Priyanka Bharti[1], Ankita Lakra[2], Abgeena Abgeena[3], Ravi Kumar Burman[4], Muskan Pandey[5], Abhishek Kumar[6]**

*[1]Jharkhand University of Technology, Data Sciences, Ranchi, India*
*Email: bpriyanka129@gmail.com*

*[2] Jharkhand University of Technology, Data Sciences, Ranchi, India*
*Email: nktlakra41@gmail.com*

*[3]Jharkhand University of Technology, Data Sciences, Ranchi, India*
*Email: abgeenakhan786@gmail.co*

*[4]Jharkhand University of Technology, Data Sciences, Ranchi, India*
*Email: raviburmanbit@gmail.com*

*[5]Swami Vivekanand University, Computer Science & Engineering, Sagar, India*
*Email: muskanpandey492@gmail.com*

*[6]St. Xavier's College, Computer Science, Ranchi, India*
*Email:abhishek28mca@gmail.com*

**ABSTRACT:**

Skin disease is a very common disease for humans. In the medical industry detecting skin disease and recognizing its type is a very challenging task. Due to the complexity of human skin texture and the visual closeness effect of the diseases, sometimes it is really difficult to detect the exact type. Therefore, it is necessary to detect and recognize the skin disease at its very first observation in a fast manner. In today's era, artificial intelligence (AI) is rapidly growing in medical fields. Different CNN Models are used for diagnostic purposes. These methods drastically improve the diagnosis process and also speed up the process. And We always choose a fast treatment in a busy increasing skin disease patient, which makes us to dependent on technology. In this paper, it has been taken 22 classes of skin disease images and in all 22 classes there are multiple images of each disease. There is a brief comparison between the multiple models of CNN. In this processes, six pre-trained deep neural network models VGG16, ResNet50, DenseNet121, EfficientNetB0, MobileNetV2 & InceptionV3 are compared with their accuracy and efficiency. In this Paper, the goal is achieving the fastest Distinction or recognition of multiple images of single disease by using AI Models. And It was observed that the accuracy of the VGG16 is satisfactory while in efficiency MobileNetV2 is more satisfactory.

**Keywords:** Skin Disease Classification, Deep Learning, Image Classification, VGG16, ResNet50, DenseNet121, EfficientNetB0, MobileNetV2, InceptionV3, Multi-Class Classification

**How to cite this article:** Priyanka Bharti, Ankita Lakra, Abgeena Abgeena, Ravi Kumar Burman, Abhishek Kumar, Muskan Pandey ,"Comparative Study of Deep Learning Models for Distinction on Multi-Class Skin Disease Image Classification". Int J Drug Deliv Technol. 2026;16(90s): 338-351; DOI: 10.25258/ijddt.16.90s.34

## 1. INTRODUCTION:

Skin is the most important and largest body part of the human. The 20 square feet of total body area of a human is covered by skin. It protects different vital organs of the human body from outside harm and also protects from microbes and the elements, regulates the body temperature, and allows the sensations of touch, heat, and cold. [10] But due to the variety of external and genetic factors, the skin may be influenced. Now a days there are multiple skin disease recognized with multiple variety of images of each disease. There are twenty-two diseases are taken in this paper. And each disease was holding multiple

images. [17][18] The name of the diseases are Acne, Actinic Keratosis, Benign tumors, Bullous, Candidiasis, Drug Eruption, Eczema, Infestations Bites, Lichen, Lupus, Moles, Psoriasis, Rosacea, Seborrh_Keratoses, Skin Cancer, Sun_Sunlight_Damage, Tinea, Nonnormally, Vascular, Vascular Tumors, Vasculitis, Vitiligo, Warts. The disease can be cure if it is diagnosis properly and recognize at its early stages. Over the past few years, artificial intelligence, machine learning, and deep learning techniques are rapidly developed in the medical field. With the physical symptoms, image analysis is also very important to detect different skin diseases. The image analysis process helps to detect skin disease with higher accuracy. In this research paper, Multiple Models in Deep Neural Network i.e. VGG16, ResNet50, DenseNet121, EfficientNetB0, MobileNetV2 & InceptionV3 are used and predict different skin diseases. Finally, a complete comparison between the all six models and observe the fastest recognition model with their accuracy and efficiency. [21, 22, 23]

## 2. LITERATURE REVIEW

Recent studies have demonstrated the effectiveness of deep learning for automated skin disease image classification. Traditional machine learning methods such as SVM, KNN, and Decision Trees generally depend on handcrafted features, which may not adequately capture the complex visual characteristics of skin lesions. [19][20] However, in contrast, Convolutional Neural Networks (CNNs) can learn automatically discriminative features such as texture, shape, color and lesion pattern directly from images. Several deep learning architectures such as VGG16, ResNet50, DenseNet, GoogleNet/Inception, MobileNet and EfficientNet have been promisingly applied to the classification of skin disease and skin lesions. Transfer learning, data augmentation, hybrid CNN-Transformer architectures and ensemble methods have also been shown to improve classification performance, especially in cases of limited or imbalanced datasets. [22][23] However, it is still difficult to distinguish between similar-looking skin diseases, as well as problems with class imbalance, image quality, dataset limitations, computational requirements and model generalization. Some comparative studies have only considered a small number of models or disease categories, pointing to the need for more extensive model comparison. Therefore, the present study focuses specifically on the image-based distinction and classification of 22 different skin disease classes, with multiple images available for each class. Multiple deep learning models will be trained and evaluated on the same dataset to provide a fair comparison and identify the most accurate and effective model for multi-class skin disease image recognition.

### 2.1 Research Gap:

Although many studies have proven the effectiveness of deep learning models for skin disease image classification, the number of disease categories considered is usually limited, or only a few individual models are evaluated in the existing research. Moreover, different studies used different datasets, preprocessing methods and evaluation settings. It is difficult to identify the best deep learning model in terms of accuracy and efficiency for multi-class distinction of skin disease. Furthermore, the challenge of distinguishing between visually similar skin diseases across a larger number of classes remains insufficiently explored. Therefore, there is a need for a standardized comparative study in which multiple deep learning models are trained and evaluated on the same dataset containing multiple images from 22 different skin disease classes under identical experimental conditions. This study addresses this gap by systematically comparing the performance of different deep learning models using appropriate evaluation metrics to identify the model that provides the most accurate and efficient distinction among the 22 skin disease classes.

### 2.2 Research Objective:

The objective of the study is to comparatively evaluate multiple deep learning models for accurate and efficient classification of skin disease images across 22 different disease classes. The main Objective are defining below:

1. To develop an image-based deep learning framework for classification of 22 different skin disease classes using multiple images from each class.
2. To implement and train multiple deep learning models under the same dataset and experimental conditions for fair performance comparison.
3. The Selected models are to evaluate and compare the classification Performance of the models using metrics such as accuracy, precision, recall, and F1-score.
4. To analyze the efficiency of the models in terms of training performance, computational requirements, and classification capability.

## 3. RESEARCH METHODOLOGY

The current study is about comparative experimental research approach to evaluate the performance of various deep learning models for automated skin disease classification. A standardized experimental framework is used to train and evaluate all selected models under identical conditions, enabling a fair comparison of their classification performance and computational efficiency.

### 3.1 Dataset Selection

The dataset used in this study is the Skin Disease Dataset developed by PacificRM and obtained from the Kaggle repository. It is the dataset which is holding labeled clinical skin disease images belonging to 22 different classes, making it suitable for multiclass image classification. It was selected due to its diversity of skin disease categories and suitability for evaluating deep learning models.

### 3.2 Preprocessing of Data:

Skin Disease and normalized before training. Images are resized to 224x224 pixel. Data augmentation techniques including rotation, flipping, zooming and brightness adjustment. These are used for increasing the model generalization and reducing overfitting. The dataset is then divided into training, validation, and testing sets while maintaining class distribution.

### 3.3. Dataset Breakdown

The dataset consists of 15,474 clinical skin disease images spanning 22 classes. It is divided into three subsets: the training set contains 11,152 images (approximately 72.07%), which are used to train the model; the validation set contains 2,776 images (approximately 17.94%), which are used to monitor model performance and tune hyperparameters; and the test set contains 1,546 unseen images (approximately 9.99%), which are used for final model evaluation. The training and validation sets together comprise 13,928 images, representing approximately 90% of the total dataset, while the remaining 10% is reserved for testing. During testing, the images are processed in batches of 32.

### 3.4 Selection of Deep Learning Models

Six pre-trained deep learning architectures—VGG16, ResNet50, DenseNet121, EfficientNetB0, MobileNetV2, and InceptionV3—are selected for comparative evaluation of clinical skin disease classification. These models represent diverse architectural approaches, feature extraction capabilities, and computational requirements. Their comparison enables a comprehensive assessment of classification performance, computational efficiency, and suitability for automated skin disease diagnosis.

### 3.5 Model Development Using Pre-Trained Models

Transfer learning is employed using ImageNet-pre-trained weights for all six deep learning architectures. The original classification layers are replaced with customized output layers comprising 22 neurons, corresponding to the skin disease classes. All models are trained under consistent experimental settings and hyperparameters to ensure a fair and reliable comparative evaluation.

### 3.6 Experimental Setup

All six deep learning models are implemented using the TensorFlow/Keras framework under consistent experimental settings to ensure a fair comparison. The input image size is set to 224 × 224 pixels, with the Adam optimizer, a learning rate of 0.001, a batch size of 32, and three training epochs. Categorical cross-entropy is used as the loss function, while the Softmax activation function generates class probabilities for the 22 skin disease categories.

### 3.7 Performance Evaluation

The models are evaluated using standard classification metrics, including accuracy, precision, recall, F1-score, sensitivity, specificity, and ROC-AUC, to assess their effectiveness in classifying the 22 skin disease categories.

### 3.8 Analysis of Computational Efficiency

Computational efficiency is assessed based on model size, trainable parameters, training time, inference time, and memory requirements to determine the suitability of each architecture for real-world deployment.

### 3.9 Comparative Analysis

All six architectures are compared based on classification performance and computational efficiency. The comparative results are used to identify the most effective and computationally suitable model for automated multi-class skin disease classification.

## 4. RESULT AND ANALYSIS

### 4.1 VGG16 Model

VGG16 is a convolutional neural network (CNN) architecture that extracts hierarchical image features, from basic edges and textures to complex patterns. In this study, ImageNet-pre-trained weights are used for transfer learning to classify 22 clinical skin disease categories. The model is trained for three epochs using a batch size of 32, the Adam optimizer, and categorical cross-entropy loss. The training set contains 11,152 images processed over 348 steps per epoch, while 2,776 images are used for validation. Training and validation accuracy and loss are monitored throughout the process. Finally, the model is evaluated on 1,546 unseen test images using accuracy, precision, recall, F1-score, and a confusion matrix. The classification results for all 22 classes are presented in Table 1.

Table 1: Class-Wise Classification Performance of VGG16 on the Skin Disease Test Dataset

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Acne | 0.5 | 0.23 | 0.32 | 65 |
| Actinic_Keratosis | 0.34 | 0.16 | 0.21 | 83 |
| Benign_tumors | 0.24 | 0.46 | 0.32 | 121 |
| Bullous | 0.17 | 0.18 | 0.18 | 55 |
| Candidiasis | 0.33 | 0.15 | 0.21 | 27 |
| DrugEruption | 0.25 | 0.28 | 0.26 | 61 |
| Eczema | 0.14 | 0.74 | 0.24 | 112 |
| Infestations_Bites | 0.09 | 0.02 | 0.03 | 60 |
| Lichen | 0 | 0 | 0 | 61 |
| Lupus | 0 | 0 | 0 | 34 |
| Moles | 0 | 0 | 0 | 40 |
| Psoriasis | 0.09 | 0.1 | 0.09 | 88 |
| Rosacea | 0.73 | 0.29 | 0.41 | 28 |
| Seborrh_Keratoses | 0.5 | 0.06 | 0.11 | 51 |
| SkinCancer | 0.26 | 0.09 | 0.13 | 77 |
| Sun_Sunlight_Damage | 0 | 0 | 0 | 34 |
| Tinea | 0.29 | 0.17 | 0.21 | 102 |
| Unknown_Normal | 0.89 | 0.85 | 0.87 | 189 |
| Vascular_Tumors | 0.25 | 0.03 | 0.06 | 60 |
| Vasculitis | 0.4 | 0.19 | 0.26 | 52 |
| Vitiligo | 0.6 | 0.68 | 0.64 | 82 |
| Warts | 0.4 | 0.03 | 0.06 | 64 |
| accuracy | 0 | 0 | 0.31 | 1546 |
| macro avg | 0.29 | 0.21 | 0.21 | 1546 |
| weighted avg | 0.34 | 0.31 | 0.28 | 1546 |

The VGG16 model achieved an overall test accuracy of 30.60% on 1,546 unseen skin disease images across 22 classes. The macro-average precision, recall, and F1-score were 0.29, 0.21, and 0.21, respectively. The weighted-average precision, recall, and F1-score were 0.34, 0.31, and 0.28, respectively. These results indicate that the model has limited classification performance under the current experimental settings, highlighting the need to investigate alternative architectures and optimization strategies.

### 4.2 ResNet50 Model

ResNet50 is a 50-layer deep convolutional neural network that employs residual (skip) connections to improve gradient flow and facilitate deep feature learning. In this study, ImageNet-pre-trained weights are used to classify 22 clinical skin disease categories through transfer learning. The dataset comprises 15,474 images, including 11,152 training images, 2,776 validation images, and 1,546 unseen test images. The model is fine-tuned for three epochs using a batch size of 32, the Adam optimizer, and categorical cross-entropy loss. Training and validation accuracy and loss are monitored to assess model performance.

The trained model is evaluated on the 1,546 unseen test images using the predict() function. The test data generator is configured with shuffle=False to maintain alignment between predicted labels and ground-truth labels. The classification report evaluates the model's performance across all 22 skin disease categories using precision, recall, F1-score, and overall accuracy. The detailed classification results are presented in Table 2.

Table 2: Class-Wise Classification Performance of ResNet50 on the Skin Disease Test Dataset

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Acne | 0.00 | 0.00 | 0.00 | 65 |
| Actinic_Keratosis | 0.03 | 0.01 | 0.02 | 83 |
| Benign_tumors | 0.00 | 0.00 | 0.00 | 121 |
| Bullous | 0.00 | 0.00 | 0.00 | 55 |
| Candidiasis | 0.03 | 0.04 | 0.03 | 27 |
| DrugEruption | 0.06 | 0.46 | 0.10 | 61 |
| Eczema | 0.00 | 0.00 | 0.00 | 112 |
| Infestations_Bites | 0.00 | 0.00 | 0.00 | 60 |
| Lichen | 0.01 | 0.02 | 0.01 | 61 |
| Lupus | 0.05 | 0.15 | 0.07 | 34 |
| Moles | 0.00 | 0.00 | 0.00 | 40 |
| Psoriasis | 0.04 | 0.09 | 0.05 | 88 |
| Rosacea | 0.00 | 0.00 | 0.00 | 28 |
| Seborrh_Keratoses | 0.00 | 0.00 | 0.00 | 51 |
| SkinCancer | 0.03 | 0.01 | 0.02 | 77 |
| Sun_Sunlight_Damage | 0.06 | 0.03 | 0.04 | 34 |
| Tinea | 0.10 | 0.02 | 0.03 | 102 |
| Unknown_Normal | 0.00 | 0.00 | 0.00 | 189 |
| Vascular_Tumors | 0.00 | 0.00 | 0.00 | 60 |
| Vasculitis | 0.00 | 0.00 | 0.00 | 52 |
| Vitiligo | 1.00 | 0.01 | 0.02 | 82 |
| Warts | 0.04 | 0.28 | 0.07 | 64 |
| accuracy | 0.00 | 0.00 | 0.04 | 1,546 |
| macro avg | 0.07 | 0.05 | 0.02 | 1,546 |
| Weighted avg | 0.07 | 0.04 | 0.02 | 1546 |

The ResNet50 model achieved an overall test accuracy of 4% across 1,546 unseen skin disease images belonging to 22 classes. The macro-average precision, recall, and F1-score were 0.07, 0.05, and 0.02, respectively. Similarly, the weighted-average precision, recall, and F1-score were 0.07, 0.04, and 0.02, respectively. These results indicate limited classification performance under the current experimental settings.

### 4.3 DenseNet121 Model

DenseNet121 is a convolutional neural network architecture that connects each layer to preceding layers within a dense block, promoting feature reuse and improving gradient propagation. This design supports efficient feature extraction and makes DenseNet121 suitable for transfer learning in clinical skin disease classification.

The dataset comprises 15,474 images across 22 skin disease classes, divided into 11,152 training images, 2,776 validation images, and 1,546 unseen test

images.

The model is trained using the skin disease dataset, with model checkpointing to save the best-performing model based on validation accuracy. Early stopping is employed to terminate training when the monitored validation metric fails to improve, helping reduce overfitting.

The trained DenseNet121 model is evaluated on the 1,546 unseen test images. Its classification performance is assessed using accuracy, precision, recall, F1-score, and class-wise support. The classification report summarizes the results across all 22 classes, as presented in Table 3.

Table 3: Class-Wise Classification Performance of DenseNet121 on the Skin Disease Test Dataset

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Acne | 0.43 | 0.65 | 0.52 | 65 |
| Actinic_Keratosis | 0.62 | 0.24 | 0.35 | 83 |
| Benign_tumors | 0.31 | 0.53 | 0.39 | 121 |
| Bullous | 0.56 | 0.27 | 0.37 | 55 |
| Candidiasis | 0.40 | 0.30 | 0.34 | 27 |
| DrugEruption | 0.54 | 0.33 | 0.41 | 61 |
| Eczema | 0.28 | 0.53 | 0.37 | 112 |
| Infestations_Bites | 0.23 | 0.25 | 0.24 | 60 |
| Lichen | 0.62 | 0.13 | 0.22 | 61 |
| Lupus | 0.18 | 0.12 | 0.14 | 34 |
| Moles | 0.26 | 0.38 | 0.31 | 40 |
| Psoriasis | 0.53 | 0.09 | 0.16 | 88 |
| Rosacea | 0.80 | 0.14 | 0.24 | 28 |
| Seborrh_Keratoses | 0.37 | 0.41 | 0.39 | 51 |
| SkinCancer | 0.43 | 0.29 | 0.34 | 77 |
| Sun_Sunlight_Damage | 0.33 | 0.03 | 0.05 | 34 |
| Tinea | 0.25 | 0.49 | 0.33 | 102 |
| Unknown_Normal | 0.96 | 0.95 | 0.95 | 189 |
| Vascular_Tumors | 0.16 | 0.32 | 0.21 | 60 |
| Vasculitis | 0.67 | 0.23 | 0.34 | 52 |
| Vitiligo | 0.87 | 0.76 | 0.81 | 82 |
| Warts | 0.53 | 0.30 | 0.38 | 64 |
| accuracy | 0.00 | 0.00 | 0.43 | 1,546 |
| macro avg | 0.47 | 0.35 | 0.36 | 1,546 |
| weighted avg | 0.51 | 0.43 | 0.42 | 1,546 |

4.4 EfficientNetB0 Model

EfficientNetB0 is a convolutional neural network architecture developed by Google that employs compound scaling to balance network depth, width, and input resolution. This approach aims to achieve effective feature extraction with improved parameter efficiency, making the architecture suitable for transfer learning in clinical skin disease classification.

The dataset comprises 15,474 clinical skin disease images across 22 classes, including 11,152 training images, 2,776 validation images, and 1,546 test images. During training, the ModelCheckpoint callback saves the best-performing model based on validation accuracy, while Early Stopping terminates training when the monitored validation metric fails to improve. During evaluation, the model generates class probability predictions using model.predict() on the evaluation dataset. Predicted labels are compared with the ground-truth labels to calculate precision, recall, F1-score, support, overall accuracy, and macro- and weighted-average metrics. The detailed classification results are presented in Table 4.

The EfficientNetB0 model achieved an overall test accuracy of 12.23% on 1,546 images across 22 skin disease classes. The macro-average precision, recall, and F1-score were 0.0056, 0.0455, and 0.0099, respectively. The weighted-average precision, recall, and F1-score were 0.0149, 0.1223, and 0.0266, respectively. These results indicate limited classification performance under the current experimental settings.

Table 4: Class-Wise Classification Performance of EfficientNetB0 on the Skin Disease Test Dataset

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Acne | 0.0000 | 0.0000 | 0.0000 | 65 |
| Actinic_Keratosis | 0.0000 | 0.0000 | 0.0000 | 83 |
| Benign_tumors | 0.0000 | 0.0000 | 0.0000 | 121 |
| Bullous | 0.0000 | 0.0000 | 0.0000 | 55 |
| Candidiasis | 0.0000 | 0.0000 | 0.0000 | 27 |
| DrugEruption | 0.0000 | 0.0000 | 0.0000 | 61 |
| Eczema | 0.0000 | 0.0000 | 0.0000 | 112 |
| Infestations_Bites | 0.0000 | 0.0000 | 0.0000 | 60 |
| Lichen | 0.0000 | 0.0000 | 0.0000 | 61 |
| Lupus | 0.0000 | 0.0000 | 0.0000 | 34 |
| Moles | 0.0000 | 0.0000 | 0.0000 | 40 |
| Psoriasis | 0.0000 | 0.0000 | 0.0000 | 88 |
| Rosacea | 0.0000 | 0.0000 | 0.0000 | 28 |
| Seborrh_Keratoses | 0.0000 | 0.0000 | 0.0000 | 51 |
| SkinCancer | 0.0000 | 0.0000 | 0.0000 | 77 |
| Sun_Sunlight_Damage | 0.0000 | 0.0000 | 0.0000 | 34 |
| Tinea | 0.0000 | 0.0000 | 0.0000 | 102 |
| Unknown_Normal | 0.1223 | 1.0000 | 0.2179 | 189 |
| Vascular_Tumors | 0.0000 | 0.0000 | 0.0000 | 60 |
| Vasculitis | 0.0000 | 0.0000 | 0.0000 | 52 |
| Vitiligo | 0.0000 | 0.0000 | 0.0000 | 82 |
| Warts | 0.0000 | 0.0000 | 0.0000 | 64 |
| accuracy | | | 0.1223 | 1,546 |
| macro avg | 0.0056 | 0.0455 | 0.0099 | 1,546 |
| weighted avg | 0.0149 | 0.1223 | 0.0266 | 1,546 |

### 4.5 MobileNetV2 Model

MobileNetV2 is an efficient convolutional neural network designed for mobile and resource-constrained applications. It uses inverted residual blocks, linear bottlenecks, and depthwise separable convolutions to reduce computational complexity while maintaining effective feature extraction. In this study, the ImageNet-pre-trained MobileNetV2 model is employed for transfer learning, with the base layers frozen during training. The dataset comprises 13,928 images across 22 skin disease classes, divided into 11,143 training images and 2,785 validation images. The model generates predictions using model.predict(), and the argmax function determines the predicted class labels. Performance is assessed using precision, recall, F1-score, and support, along with macro- and weighted-average metrics. The detailed classification report is presented in Table 5.

The MobileNetV2 model achieved an overall accuracy of 38% on 2,785 validation images. The macro-average precision, recall, and F1-score were 0.333, 0.31, and 0.30, respectively, while the weighted-average precision, recall, and F1-score were 0.37, 0.38, and 0.36, respectively. These results indicate moderate classification performance under the reported experimental settings.

Table 5: Class-Wise Classification Performance of MobileNetV2 on the Skin Disease Evaluation Dataset

| Class / Condition | precision | recall | f1-score | support |
|---|---|---|---|---|
| Acne | 0.47 | 0.64 | 0.54 | 118 |
| Actinic_Keratosis | 0.36 | 0.48 | 0.41 | 137 |
| Benign_tumors | 0.29 | 0.35 | 0.32 | 237 |
| Bullous | 0.39 | 0.21 | 0.28 | 103 |
| Candidiasis | 0.29 | 0.08 | 0.13 | 49 |
| DrugEruption | 0.40 | 0.25 | 0.31 | 111 |
| Eczema | 0.36 | 0.45 | 0.40 | 217 |
| Infestations_Bites | 0.22 | 0.12 | 0.16 | 98 |
| Lichen | 0.13 | 0.09 | 0.10 | 102 |
| Lupus | 0.11 | 0.02 | 0.03 | 63 |
| Moles | 0.19 | 0.27 | 0.22 | 63 |
| Psoriasis | 0.26 | 0.16 | 0.20 | 171 |
| Rosacea | 0.31 | 0.43 | 0.36 | 46 |
| Seborrh_Keratoses | 0.51 | 0.28 | 0.36 | 104 |
| SkinCancer | 0.35 | 0.29 | 0.32 | 136 |
| Sun_Sunlight_Damage | 0.20 | 0.04 | 0.06 | 54 |
| Tinea | 0.22 | 0.45 | 0.29 | 181 |
| Unknown_Normal | 0.81 | 0.89 | 0.85 | 325 |
| Vascular_Tumors | 0.16 | 0.15 | 0.15 | 124 |
| Vasculitis | 0.30 | 0.19 | 0.23 | 97 |
| Vitiligo | 0.60 | 0.65 | 0.63 | 138 |
| Warts | 0.30 | 0.28 | 0.29 | 111 |
| accuracy | 0.00 | 0.00 | 0.38 | 2,785 |
| macro avg | 0.33 | 0.31 | 0.30 | 2,785 |
| weighted avg | 0.37 | 0.38 | 0.36 | 2,785 |

### 4.6 InceptionV3 Model

InceptionV3 is a deep convolutional neural network that employs parallel convolutional operations to extract image features at multiple spatial scales. Its architecture balances feature representation and computational efficiency, making it suitable for transfer learning in clinical skin disease classification. In this study, the ImageNet-pre-trained InceptionV3 base is used with its layers frozen during training. The dataset contains 13,928 images across 22 skin disease classes, comprising 11,143 training images and 2,785 validation images.

During evaluation, model.predict() generates class probability predictions, and argmax identifies the predicted class labels. The model is assessed using precision, recall, F1-score, and support for each class, together with macro- and weighted-average metrics. The InceptionV3 model achieved an overall test accuracy of 39.07% on 1,546 unseen skin disease images across 22 classes. The macro-average precision, recall, and F1-score were 0.3230, 0.3088, and 0.2881, respectively. The weighted-average precision, recall, and F1-score were 0.3740, 0.3907, and 0.3546, respectively. These results provide an overall assessment of the model's classification performance, with the detailed class-wise results presented in Table 6.

Table 6: Class-Wise Classification Performance of InceptionV3 on the Skin Disease Test Dataset

| Class / Condition | precision | recall | f1-score | support |
|---|---|---|---|---|
| Acne | 0.4245 | 0.6923 | 0.5263 | 65 |
| Actinic_Keratosis | 0.3402 | 0.3976 | 0.3667 | 83 |
| Benign_tumors | 0.2903 | 0.4463 | 0.3518 | 121 |
| Bullous | 0.3721 | 0.2909 | 0.3265 | 55 |
| Candidiasis | 0.0000 | 0.0000 | 0.0000 | 27 |
| DrugEruption | 0.5625 | 0.1475 | 0.2338 | 61 |
| Eczema | 0.2557 | 0.5000 | 0.3384 | 112 |
| Infestations_Bites | 0.4000 | 0.0333 | 0.0615 | 60 |
| Lichen | 0.3750 | 0.0984 | 0.1558 | 61 |
| Lupus | 0.1250 | 0.1471 | 0.1351 | 34 |
| Moles | 0.0000 | 0.0000 | 0.0000 | 40 |
| Psoriasis | 0.2812 | 0.1023 | 0.1500 | 88 |
| Rosacea | 0.3659 | 0.5357 | 0.4348 | 28 |
| Seborrh_Keratoses | 0.4615 | 0.3529 | 0.4000 | 51 |
| SkinCancer | 0.2658 | 0.2727 | 0.2692 | 77 |
| Sun_Sunlight_Damage | 0.0000 | 0.0000 | 0.0000 | 34 |
| Tinea | 0.2294 | 0.3824 | 0.2868 | 102 |
| Unknown_Normal | 0.7468 | 0.9365 | 0.8310 | 189 |
| Vascular_Tumors | 0.1818 | 0.0333 | 0.0563 | 60 |
| Vasculitis | 0.4583 | 0.4231 | 0.4400 | 52 |
| Vitiligo | 0.6849 | 0.6098 | 0.6452 | 82 |
| Unknown_Normal | 0.1223 | 1.0000 | 0.2179 | 189 |
| Vascular_Tumors | 0.0000 | 0.0000 | 0.0000 | 60 |
| Vasculitis | 0.0000 | 0.0000 | 0.0000 | 52 |
| Vitiligo | 0.0000 | 0.0000 | 0.0000 | 82 |
| Warts | 0.0000 | 0.0000 | 0.0000 | 64 |
| accuracy | 0.000 | 0.000 | 0.1223 | 1,546 |
| macro avg | 0.0056 | 0.0455 | 0.0099 | 1,546 |
| weighted avg | 0.0149 | 0.1223 | 0.0266 | 1,546 |

### 4.7 Comparative Analysis of Classification Performance

The six pre-trained deep learning architectures are compared based on their reported test accuracy, precision, recall, and F1-score to identify their effectiveness in multi-class skin disease image classification. Table 7 summarizes the reported classification accuracy of each model.

Table 7: Comparative Performance of Six Deep Learning Models Based on Classification Metrics

| Model Name | Recall | | Precision | | F1-Score | | Supporting | |
|---|---|---|---|---|---|---|---|---|
| | Macro_Avg | Weighted Avg | Macro_Avg | Weighted Avg | Macro_Avg | Weighted Avg | Macro_Avg | Weighted Avg |
| VGG16 | 0.29 | 34 | 0.21 | 31 | 0.21 | 0.28 | 1546 | 1546 |
| ResNet50 | 0.02 | 0.03 | 0.04 | 0.04 | 0.02 | 0.02 | 1546 | 1546 |
| DenseNet121 | 0.47 | 0.51 | 0.35 | 0.43 | 0.36 | 0.42 | 1546 | 1546 |
| EfficientNetB0 | 0.0056 | 0.0149 | 0.0455 | 0.1223 | 0.0099 | 0.0266 | 1546 | 1546 |
| MobileNetV2 | 0.33 | 0.37 | 0.31 | 0.38 | 0.3 | 0.36 | 2785 | 2785 |
| InceptionV3 | 0.323 | 0.374 | 0.3088 | 0.3907 | 0.2881 | 0.3546 | 1546 | 1546 |

The comparative results indicate that DenseNet121 achieved the highest reported classification accuracy of 43%, followed by InceptionV3 at 39.07% and MobileNetV2 at 38.17%. VGG16 achieved 30.60%, while EfficientNetB0 and ResNet50 recorded substantially lower accuracies of 12.23% and 4.33%, respectively. These findings suggest that DenseNet121 provides the strongest classification performance among the evaluated architectures under the reported experimental settings. However, the differences in model performance also indicate that architectural design and implementation choices may influence classification effectiveness across the 22 skin disease categories.

Figure 1 illustrates the comparative classification accuracy of six pre-trained deep learning models for multiclass skin disease image classification. The comparison highlights differences in model performance, with DenseNet121 achieving the highest reported accuracy, followed by InceptionV3 and MobileNetV2.

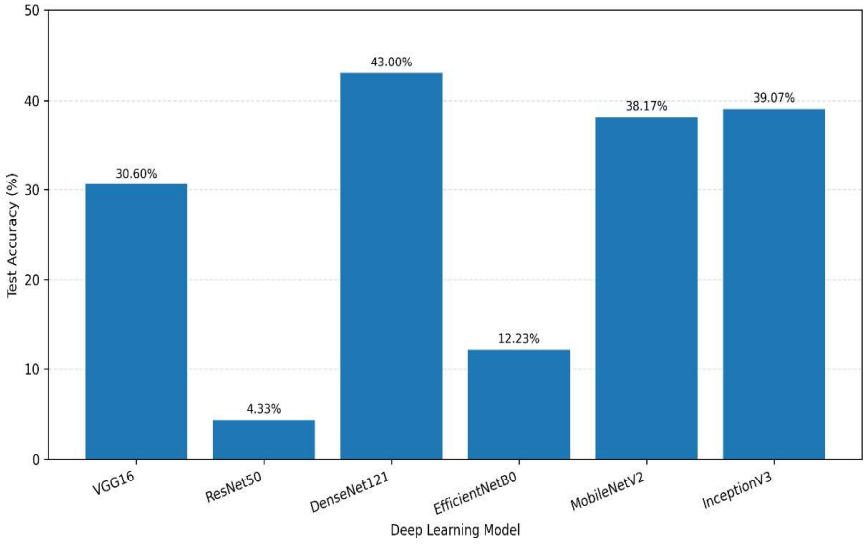


Figure 1: Comparative Classification Accuracy of Six Pre-trained Deep Learning Models for Multiclass Skin Disease Image Classification

### 4.8 Computational Efficiency Analysis

Computational efficiency is evaluated using the total prediction time and average inference time per image. These measures help assess the suitability of each architecture for applications requiring rapid image classification.

Table 8: Comparative Analysis of Computational Efficiency Based on Prediction Time

| Model | Images Evaluated | Total Prediction Time (s) | Average Time per Image (s) |
|---|---|---|---|
| VGG16 | 1,546 | 825 | 0.534 |
| ResNet50 | 1,546 | 273 | 0.177 |
| DenseNet121 | 1,546 | 510 | 0.330 |
| EfficientNetB0 | 1,546 | 411 | 0.266 |
| MobileNetV2 | 2,785 | 127 | 0.046 |
| InceptionV3 | 1,546 | 698 | 0.451 |

The results indicate that MobileNetV2 recorded the shortest total prediction time of 127 seconds and the lowest average inference time, approximately 45.6 milliseconds per image. ResNet50 followed with an average inference time of approximately 176.6 milliseconds per image. EfficientNetB0 and DenseNet121 required approximately 265.8 and 329.9 milliseconds per image, respectively, while InceptionV3 and VGG16 recorded approximately 451.5 and 533.6 milliseconds per image.

These findings suggest that MobileNetV2 offers promising computational efficiency, making it a potential candidate for resource-constrained image-classification applications. Nevertheless, MobileNetV2 was evaluated on 2,785 images, whereas the other models were evaluated on 1,546 images. Therefore, a definitive comparison of computational efficiency requires all models to be benchmarked using the same test images, hardware, batch size, and timing procedure.

## 5. DISCUSSION

The comparative study demonstrates differences in classification performance and prediction speed among the six pre-trained deep learning architectures. DenseNet121 achieved the highest reported accuracy of 43%, indicating comparatively stronger classification performance on the evaluated dataset. InceptionV3 and MobileNetV2 followed with reported accuracies of 39.07% and 38.17%, respectively. In contrast, ResNet50 and EfficientNetB0 exhibited substantially lower reported accuracies, suggesting that their current implementations may require further optimization.

From a computational perspective, MobileNetV2 demonstrated the shortest reported prediction time and lowest average inference time per image. This finding supports its potential for applications where computational resources and response time are important considerations. DenseNet121, on the other hand, offers the highest reported accuracy among the evaluated models, making it a candidate for further investigation where classification performance is prioritized.

Overall, the findings highlight a trade-off between classification accuracy and computational efficiency. DenseNet121 is the strongest candidate based on the reported accuracy, whereas MobileNetV2 is the most promising candidate for rapid inference. However, neither model should be considered definitively superior until the evaluation conditions and reported metrics are made consistent across all six architectures.

## 6. CONCLUSION

This study compared six pre-trained convolutional neural network architectures for classifying 22 skin disease categories. Based on the reported results, DenseNet121 achieved the highest classification accuracy at 43%, while MobileNetV2 demonstrated the fastest reported inference, at approximately 45.6 milliseconds per image. The results illustrate the importance of considering both classification performance and computational efficiency when selecting a deep learning architecture.

Further experimentation using consistent data splits, standardized preprocessing, identical test samples, and repeated performance measurements is necessary to establish a reliable model ranking. In addition, external validation and clinical assessment would be required before considering any model for practical diagnostic use.

**REFERENCES:**


1. Bose, P., Bandyopadhyay, S. K., Bhaumik, A., & Poddar, S. (2021). Skin disease detection: Machine learning vs deep learning. *Preprints*. doi:10.20944/preprints202109.0209.v1 Skin Disease Detection: Machine Learning vs Deep Learning[v1] | Preprints.org
2. 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. https://doi.org/10.33545/27076636.2022.v3.i1b.53
3. Bhosale, S. (2022). *Comparison of deep learning and machine learning models and frameworks for skin lesion classification*. arXiv. https://doi.org/10.48550/arXiv.2207.12715
4. Li, L. F., Wang, X., Hu, W. J., Xiong, N. N., Du, Y. X., & Li, B. S. (2020). Deep learning in skin disease image recognition: A review. *IEEE Access*, *8*, 208264-208280. https://doi.org/10.1109/access.2020.3037258
5. Senthilkumar, C., Sindhu, C., Vadivu, G., & Neethirajan, S. (2024). Early detection of lumpy skin disease in cattle using deep learning—A comparative analysis of pretrained models. *Veterinary Sciences*, *11*(10), 510. https://doi.org/10.3390/vetsci11100510
6. Sudheesh, K. V., Kumara, B., Puttegowda, K., Kavitha, R. J., Manasa, S., Divya, B. N., & Sunikumar, D. S. (2026). Early detection of skin diseases using deep learning approaches. In *Information and Communication Systems* (pp. 126–132). Information and Communication Systems
7. Liu, R., Chen, Z., Yao, G., & Zhang, P. (2025). Exploring the challenge and value of deep learning in automated skin disease diagnosis. *arXiv*. https://doi.org/10.48550/arxiv.2510.03869
8. S. Ahmed, K. M. S. Sakib, and U. Hany, "Hybrid deep learning architecture for skin disease classification," *Franklin Open*, vol. 15, p. 100556, Feb. 2026, DOI: 10.1016/j.fraope.2026.100556.
9. Maram, P., Ramasubramanian, P., & Venkatasubbu, P. (2025). Hybrid explainable deep learning framework for skin disease detection and future risk prediction using Sakaguchi function with CNN-transformer models. *Results in Engineering*, *29*, Article 108489. https://doi.org/10.1016/j.rineng.2025.108489
10. Devi, T. B., & Zamzachin, G. (2006). Pattern of skin diseases in Imphal. *Indian Journal*
11. Li, H., Pan, Y., Zhao, J., & Zhang, L. (2020). Skin disease diagnosis with deep learning: A review. *arXiv*. https://doi.org/10.48550/arXiv.2011.05627
12. Kassem, M. A., Hosny, K. M., & Fouad, M. M. (2020). Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning. *IEEE Access*, *8*, 114822–114832. https://doi.org/10.1109/ACCESS.2020.3003 890
13. Bi, L., Kim, J., Ahn, E., & Feng, D. (2017). Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks.*arXiv*. https://doi.org/10.48550/arxiv.1703.04197.
14. Khan, M. A., Sharif, M., Akram, T., Damaševičius, R., & Maskeliūnas, R. (2021). Skin lesion segmentation and multiclass classification using deep learning features and improved moth flame optimization. *Diagnostics*, *11*(5), 811. https://doi.org/10.3390/diagnostics1105081 1
15. Celebi, M. E., Kingravi, H. A., Uddin, B., Iyatomi, H., Aslandogan, Y. A., Stoecker, W. V., & Moss, R. H. (2007). A methodological approach to the classification of dermoscopy images. *Computer Methods and Programs in Biomedicine*, 87(1), DOI.10.1016/j.compmedimag.2007.01.003
16. A. Kumar Baitha, A. Kumar, and R. Murmu, "Disease Identification for Lac Crops using CNN Model," 2024 International BIT Conference (BITCON), pp.1–6, Dec. 2024, https://doi.org/10.1109/bitcon63716.2024.10985687.
17. S. Jindal, G. Gupta, and R. Misra, "The Impact of Neuromorphic Computing on Brain-Inspired Robotics: A Novel Architectural Paradigm," 2025 3rd International Conference on Communication, Security, and Artificial Intelligence (ICCSAI), pp. 1888–1892, Apr. 2025, doi: https://doi.org/10.1109/iccsai64074.2025.11064198
18. G. Gupta and E. R. Misra, "Keren-FFTGAN: A Novel Technique for Super-resolution Image Reconstruction Using Keren and FFT Image Registration with Generative Adversarial Networks," Research Square (Research Square), Mar. 2024, doi: https://doi.org/10.21203/rs.3.rs-4136145/v1.

19. V. Kumar, R. K. Burman, A. Kumar, N. Kumar, M. S. Alam, and P. Kumar, "Explainable Artificial Intelligence(XAI) in Cardiac Disease: Using SHAP Technique," Communications in computer and information science, pp. 105–118, Jul. 2025, doi: https://doi.org/10.1007/978-3-031-92041-7_9.
20. Munda, B., Kumar, A., Burman, R.K., Aisiri, A.P., Singh, R. (2026). Modeling and Forecasting the Impact of the Swachh Bharat Mission on Rural Women's Health in Jharkhand: A Machine Learning Approach. In: Bhattacharya, A., Dutta, S., Yang, XS., Goon, S. (eds) Innovations in Data Analytics. ICIDA 2025. Lecture Notes in Networks and Systems, vol 2008. Springer, Cham. https://doi.org/10.1007/978-3-032-27711-4_25.
21. P. Tiwari, R. K. Burman, and A. Kumar, "Compare and evaluate AI models for automatically classifying and categorizing URIs," Journal of Computers, Mechanical and Management, vol. 4, no. 2, pp. 24–29, Apr. 2025, doi: https://doi.org/10.57159/jcmm.4.2.25196
22. Pathan, S., Prabhu, K. G., & Siddalingaswamy, P. C. (2018). Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—A review. Biomedical Signal Processing and Control, 39, 237–262. https://doi.org/10.1016/j.bspc.2017.07.010
23. Khan, M. Q., et al. (2019). Classification of melanoma and nevus in digital images for diagnosis of skin cancer. IEEE Access, 7, 90132–90144. https://doi.org/10.1109/ACCESS.2019.2926837
24. Maiti, A., Chatterjee, B., Ashour, A. S., & Dey, N. (2020). Computer-aided diagnosis of melanoma: A review of existing knowledge and strategies. Current Medical Imaging, 16(7), 835–854. https://doi.org/10.2174/1573405615666191210104141
25. De Bruijne, M. (2016). Machine learning approaches in medical image analysis: From detection to diagnosis. Medical Image Analysis, 33, 94–97. https://doi.org/10.1016/j.media.2016.06.032
26. Rotemberg, V., et al. (2021). A patient-centric dataset of images and metadata for identifying melanomas using clinical context. Scientific Data, 8, Article 34. https://doi.org/10.1038/s41597-021-00815-z
27. Nachbar, F., et al. (1994). The ABCD rule of dermatoscopy. Journal of the American Academy of Dermatology, 30(4), 551–559. https://doi.org/10.1016/S0190-9622(94)70061-3
28. Ballerini, L., Fisher, R., Aldridge, B., & Rees, J. (2013). A color and texture based hierarchical K-NN approach to the classification of non-melanoma skin lesions. Lecture Notes in Computational Vision and Biomechanics, 6, 63–86. https://doi.org/10.1007/978-94-007-5389-1_4
29. Vasconcelos, M. J. M., Rosado, L., & Ferreira, M. (2015). A new color assessment methodology using cluster-based features for skin lesion analysis. In 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) (pp. 373–378). https://doi.org/10.1109/MIPRO.2015.7160298
30. Painuli, D., Bhardwaj, S., & Köse, U. (2022). Recent advancement in cancer diagnosis using machine learning and deep learning techniques: A comprehensive review. Computers in Biology and Medicine, 146, Article 105580. https://doi.org/10.1016/j.compbiomed.2022.105580
31. Ibrahim, A. T., Abdullahi, M., Kana, A. F. D., Mohammed, M. T., & Hassan, I. H. (2024). Categorical classification of skin cancer using a weighted ensemble of transfer learning with test time augmentation. Data Science and Management. https://doi.org/10.1016/j.dsm.2024.10.002
32. Ahmad, I., & Alqurashi, F. (2024). Early cancer detection using deep learning and medical imaging: A survey. Critical Reviews in Oncology/Hematology, 204, Article 104528. https://doi.org/10.1016/j.critrevonc.2024.104528
33. Huang, Q., Ding, H., & Sheykhahmad, F. R. (2023). A skin cancer diagnosis system for dermoscopy images according to deep training and metaheuristics. Biomedical Signal Processing and Control, 83, Article 104705. https://doi.org/10.1016/j.bspc.2023.104705
34. Adegun, S., & Viriri, T. (2021). FCN-based skin lesion segmentation model with probabilistic graphical inference. Biocybernetics and Biomedical Engineering, 41(2), 676–691. https://doi.org/10.1016/j.bbe.2021.04.008
35. Deepa, M., Arulmurugan, K., & Kandaswamy, B. (2023). Adaptive boundary-aware transformer for skin lesion segmentation with hybrid optimization. Biomedical Signal Processing and Control, 81, Article 104413. https://doi.org/10.1016/j.bspc.2023.104413

36. Ahmad, M., Khan, S., & Hussain, T. (2024). A dual-phase pipeline with Deeplabv3+ and vision transformer for robust skin lesion analysis. Biomedical Signal Processing and Control, 85, Article 104895. https://doi.org/10.1016/j.bspc.2024.104895

37. Kumar, R., Kumar, S. S., & Panda, N. (2022). Hybrid U-RP-net segmentation and aquila whale optimized SqueezeNet for skin cancer detection. Biocybernetics and Biomedical Engineering, 42(3), 851–865. https://doi.org/10.1016/j.bbe.2022.06.005

38. Albraikan, B., Basha, A. H., & El-dosuky, M. A. (2022). ADL-MDC: Automated deep learning-based melanoma detection and classification model. Computers in Biology and Medicine, 149, Article 106125. https://doi.org/10.1016/j.compbiomed.2022.106125

39. Rasheed, J., Anwar, S., & Jamil, N. (2022). Hybrid deep learning model for automatic eczema classification. Computers in Biology and Medicine, 149, Article 106210. https://doi.org/10.1016/j.compbiomed.2022.106210

40. Huo, Y., Zhang, X., & Li, W. (2023). HiFuse: Hierarchical multi-scale feature fusion network for skin lesion analysis. Computers in Biology and Medicine, 157, Article 106787. https://doi.org/10.1016/j.compbiomed.2023.106787

41. Jasil, A., Thomas, S., & Bindu, P. (2023). A hybrid DenseNet-ResNet model with multi-task learning for skin lesion classification and segmentation. Computers in Biology and Medicine, 158, Article 106857. https://doi.org/10.1016/j.compbiomed.2023.106857.