# Derma AI: Automated Skin Condition Classification Using Deep Learning

Aman Kudiyal[1] Abhishek[2] Parveen Kumar[3] YashRaj Ojha[4]
aman21128@example.com, abhishek21121@example.com, parveen21177@example.com, yashraj21219@example.com

November 29, 2024

## Abstract

Skin diseases pose significant global health challenges, often requiring timely and accurate diagnosis. This paper presents *Derma AI*, an automated skin condition classification system leveraging deep learning techniques to assist in dermatological diagnostics. We explore various Convolutional Neural Network (CNN) architectures, including VGG16, ResNet50, MobileNetV2, and EfficientNetB0, along with traditional machine learning models like Support Vector Classifier (SVC), K-Nearest Neighbors (KNN), and Random Forest. Using a dataset of ten skin condition classes, we apply data preprocessing and augmentation techniques to address class imbalance and improve generalization. ResNet50 achieved the highest validation accuracy of 68.38%, highlighting the efficacy of deeper architectures. Challenges such as image quality variability and underrepresentation of darker skin tones are discussed. Future work will focus on dataset expansion, model refinement, and incorporating multimodal data to enhance diagnostic accuracy.

## 1 Introduction

Skin diseases range from common conditions like acne and eczema to severe disorders such as melanoma, necessitating accurate diagnosis for effective treatment. Traditional diagnostic methods rely heavily on visual examination by dermatologists, which can be subjective, time-consuming, and inconsistent, especially in regions with limited access to specialized healthcare. The increasing global prevalence of skin diseases highlights the need for innovative solutions that can assist in timely and accurate diagnosis.

*Derma AI* aims to address these challenges by automating skin condition classification using deep learning techniques. By leveraging Convolutional Neural Networks (CNNs), which have demonstrated exceptional performance in image classification tasks, we seek to develop a tool that can aid healthcare professionals in diagnosing a wide range of skin conditions. This system has the potential to improve diagnostic accuracy, reduce workload for dermatologists, and increase accessibility to dermatological care in underserved areas.

We explore multiple CNN architectures, including VGG16, ResNet50, MobileNetV2, and EfficientNetB0, to identify the most effective model for this task. Additionally, we compare these deep learning models with traditional machine learning approaches like Support Vector Classifier (SVC), K-Nearest Neighbors (KNN), and Random Forest to evaluate their performance differences. Techniques such as transfer learning, data augmentation, and hyperparameter tuning are employed to enhance model performance and generalization.

## 2 Related Work

Artificial Intelligence (AI) applications in dermatology have shown significant promise but also face challenges such as dataset bias and underrepresentation of skin of color (SOC). The work by Florent et al. [1] highlights that most AI models in dermatology are trained on datasets predominantly containing images of lighter skin tones, leading to biases and reduced diagnostic accuracy for individuals with darker skin. This underrepresentation can exacerbate health disparities and limit the effectiveness of AI tools across diverse populations.

Another study by Yap et al. [2] demonstrates the benefits of multimodal data fusion in enhancing classification accuracy for skin lesions. By integrating dermatoscopic and macroscopic images with patient metadata (e.g., age, gender, lesion location), the authors improved the model's ability to classify complex conditions like melanoma. This approach underscores the potential of combining various data sources to improve AI diagnostic performance.

These studies inform our approach by emphasizing the importance of dataset diversity and considering multimodal inputs. Addressing biases and incorporating additional data types can enhance model robustness and ensure equitable performance across different skin tones and conditions.

# 3 Dataset and Preprocessing

## 3.1 Dataset Overview

The dataset comprises dermatological images representing ten skin condition classes, including acne, eczema, psoriasis, and melanoma. Images are sourced from publicly available repositories, providing a diverse set of skin types and conditions. However, the dataset presents challenges such as class imbalance and variability in image quality, resolution, and lighting conditions.

## 3.2 Preprocessing Techniques

To prepare the dataset for model training and address the aforementioned challenges, we implement the following preprocessing steps:

- **Resizing**: All images are resized to $224 \times 224$ pixels to ensure consistency and compatibility with pre-trained CNN architectures.
- **Normalization**: Pixel values are normalized to the [0, 1] range by dividing by 255, which aids in faster convergence during training.
- **Grayscale Conversion**: For specific experiments requiring single-channel inputs, images are converted to grayscale, simplifying the data by reducing it to intensity values.
- **Histogram Equalization**: Contrast is enhanced by redistributing intensity values, improving the visibility of features in poorly lit or low-contrast images.
- **Data Augmentation**: Techniques such as random rotations, horizontal flipping, zooming, and shifting are applied to increase data diversity and mitigate overfitting.
- **Class Balancing**: Underrepresented classes are augmented to balance the number of samples per class, reducing model bias towards majority classes.

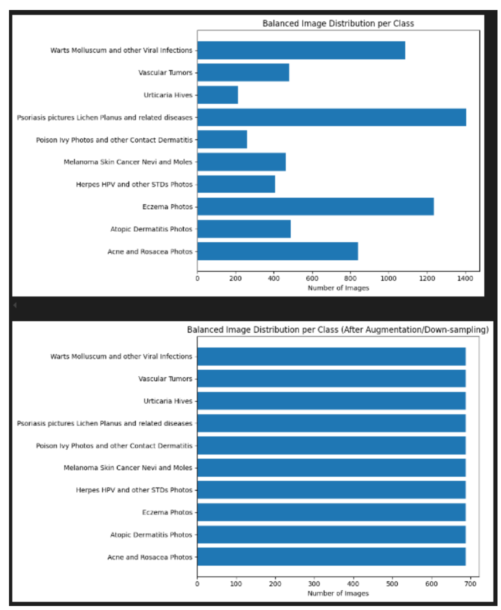


Figure 1: Class distribution before and after balancing the dataset.

### 3.3 Data Splitting

The dataset is split into training, validation, and test sets using an 80-10-10 split, ensuring that each set contains a representative distribution of classes. Stratified sampling is used to maintain class proportions across splits.

## 4 Methodology

### 4.1 Model Architectures

We explore both deep learning and traditional machine learning models:

- **Basic CNN**: A custom CNN with convolutional, pooling, and fully connected layers, serving as a baseline.
- **Pre-trained Models**: We fine-tune pre-trained models including VGG16, ResNet50, MobileNetV2, and EfficientNetB0 to leverage learned features from large datasets like ImageNet.
- **Traditional Models**: SVC, KNN, and Random Forest are trained using features extracted via Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP).

### 4.2 Training and Evaluation

Models are trained using the Adam optimizer with categorical cross-entropy loss for multi-class classification. Early stopping and learning rate reduction on plateau are implemented to prevent overfitting and improve convergence. Evaluation metrics include accuracy, precision, recall, F1-score, and ROC AUC where applicable.

### 4.3 Hyperparameter Tuning

Hyperparameters such as learning rate, batch size, and number of epochs are tuned using the validation set. For pre-trained models, different layers are unfrozen to fine-tune feature extraction capabilities. For the Random Forest classifier, the number of trees, maximum depth, and minimum samples split are optimized.

## 5 Results and Discussion

### 5.1 Model Performance

| **Model** | **Validation Accuracy** |
|---|---|
| Basic CNN | 27.48% |
| VGG16 | 38.64% |
| MobileNetV2 | 16.01% |
| EfficientNetB0 | 53.98% |
| ResNet50 | **68.38%** |
| SVC | 33.87% |
| KNN | 19.70% |
| Random Forest | 33.72% |

Table 1: Validation accuracy of different models.

ResNet50 outperformed other models, achieving a validation accuracy of 68.38%. EfficientNetB0 also showed promising results with 53.98% accuracy, benefiting from its efficient scaling of network dimensions. The Random Forest classifier achieved an accuracy of 33.72%, comparable to SVC but still lagging behind deep learning models.

### 5.2 Analysis of Results

The superior performance of ResNet50 suggests that deeper models with residual connections are more effective for skin condition classification. EfficientNetB0's compound scaling also proved beneficial. The Random Forest classifier performed better than KNN but was similar to SVC, indicating that while ensemble methods capture more complex patterns than simple classifiers, they may not match the feature extraction capabilities of deep learning models.

The Basic CNN and MobileNetV2 models showed lower accuracy, possibly due to insufficient depth or capacity to capture complex dermatological features. The discrepancy between training and validation accuracy in some models indicates potential overfitting or underfitting.

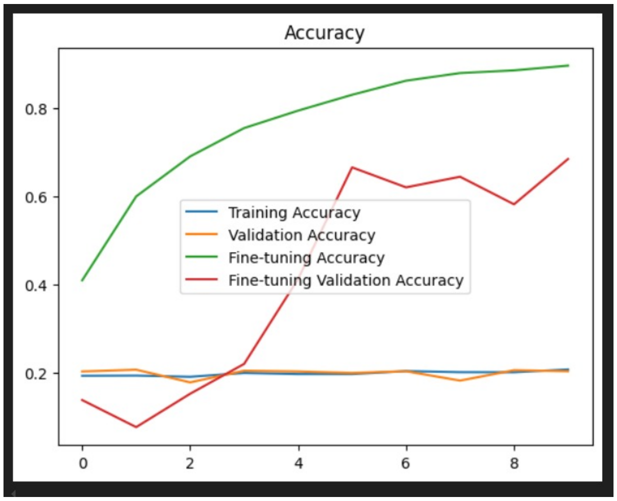


Figure 2: Training and fine-tuning accuracy for ResNet50. Fine-tuning significantly boosts validation accuracy over the epochs.

Key challenges encountered include:

- **Class Imbalance**: Despite augmentation, some classes remained underrepresented, affecting the model's ability to generalize across all conditions.
- **Image Variability**: Variations in lighting, resolution, and skin tone introduced noise, making it difficult for models to learn consistent features.
- **Underrepresentation of SOC**: Limited images of darker skin tones may introduce bias, reducing diagnostic accuracy for SOC populations [1].

### 5.3 Limitations

Our study is limited by the size and diversity of the dataset. Additionally, the lack of patient metadata prevents the exploration of multimodal approaches, which have shown to improve classification performance [2].

## 6 Conclusion

*Derma AI* demonstrates the potential of deep learning in automating skin condition classification. ResNet50 achieved the highest validation accuracy, highlighting the effectiveness of deeper architectures for this task. EfficientNetB0 also showed strong performance, indicating that efficient scaling can be beneficial. The Random Forest classifier provided results comparable to SVC but fell short compared to deep learning approaches.

Challenges such as class imbalance, image variability, and underrepresentation of SOC need to be addressed to improve model robustness and fairness. Future work will focus on expanding the dataset with more diverse images, including those representing darker skin tones. Incorporating patient metadata and exploring multimodal approaches may enhance diagnostic accuracy. Additionally, implementing advanced data augmentation and regularization techniques could further improve model generalization.

## Acknowledgments

We thank all team members for their contributions:

- **Aman Kudiyal**: Model training and evaluation.
- **Abhishek**: Model training and evaluation.
- **Parveen Kumar**: Data preprocessing, augmentation, and balancing.
- **YashRaj Ojha**: Visualization and documentation.

## References

[1] R. Florent, B. Fardman, A. Podwojniak, *et al.* Artificial intelligence in dermatology: advancements and challenges in skin of color. *International Journal of Dermatology*, 63:455–461, 2024.

[2] J. Yap, W. Yolland, and P. Tschandl. Multimodal skin lesion classification using deep learning. *Experimental Dermatology*, 27(11):1261–1267, 2018.