

SKIN DISEASE DETECTION USING VISION TRANSFORMERS

Real-Time Classification for Accessible Dermatological Care

Final Project Report

MSDSP 462 - Computer Vision

Abhigna Mallepally, Edwin Daniels

March 17, 2025

**Abstract**

Accurate and early diagnosis of skin diseases is critical particularly for patients that do not have immediate access to a dermatologist. Therefore, our project aims to develop a robust deep learning model to automatically classify dermatological images into clinically relevant categories. By leveraging state-of-the-art transformer-based architectures—specifically, a fine-tuned variant of Google’s VIT-base model—we improve diagnostic efficiency and accuracy. The goal is to advance from a research prototype toward a Minimum Viable Product (MVP) that can be deployed in clinical settings.

**Data Profile and Exploration**

The DermNet dataset comprises thousands of skin images (training and testing directories) stored in multiple formats (JPG, JPEG, AVIF) and organized into 23 disease categories like Psoriasis, Keratoses, Benign Tumors, Fungal Infections, Eczema, and Actinic Keratosis, Basal Cell Carcinoma, Malignant Lesions and more. A relatively balanced distribution is observed among the top five classes whereas the remaining 18 classes are not equally represented. Further, augmentation techniques (e.g., RandomResizedCrop and RandomHorizontalFlip) have improved image variability.

The dataset is moderately clean; however, occasionally corrupted images (e.g., unreadable files) were pre-filtered. We extracted the cleaned 15,557 images from Hugging Face via the `load\_dataset` API and split them across training, validation, and testing sets.

**Methodology**

**Data Pre-processing :** After loading the images from DermNet HuggingFace dataset, extensive augmentation (random crops, horizontal flips, brightness/contrast adjustments) is applied using both Hugging Face’s AutoImageProcessor and/or torchvision transforms,

**Model Implementation:** Initial experiments with Facebook’s DINOv2 model have achieved promising training accuracy, but faced challenges with generalization and data quality. Additionally, we performed transfer learning using MobileNet and YOLOv8 models which achieved low validation accuracies despite performance optimization and hyperparameter tuning.

The benchmarking accuracy was achieved using the fine-tuned Google-based transformer model,“vit-base-patch16-224-in21k”. We employed PyTorch and HuggingFace’s Transformers library for training this fine-tuned model on DermNet dataset for skin condition classification. Training is performed on Google Colab with A100 GPU acceleration, managed through custom PyTorch training loops to optimize performance. Initially, transfer learning techniques were employed by freezing lower-level layers initially, then gradually fine-tuning the entire network, however they did not achieve the accuracy of the fine-tuned models. Optimization techniques like regularization (dropout, weight decay), learning rate scheduling, and gradient accumulation are used to mitigate overfitting and improve generalization performance. The trained model is finally converted to an edge inference framework, ONNX, and deployed onto a Streamlit app for predictions using API inference, enabling real-time diagnostic support.

**Model Evaluation and Results**

Metrics including accuracy, precision, recall, and F1-score are used to evaluate the clinical relevance of predictions. Misclassifications are further analyzed using confusion matrix and classification reports. The fine-tuned Google ViT-base model outperformed MobileNetV2, the fine-tuned DINOv2, and YOLO v8 across all metrics.

**Key Findings**

The Google ViT-base model performed exceptionally well, achieving >90% accuracy on test images with high precision and recall across most skin disease classes. However, some misclassifications occur in visually similar conditions like Eczema, and Pigmentation disorders. It achieved high training accuracy (up to 96%), indicating effective feature learning on the training data and validation accuracy shows good model generalisation (±96%). An additional observation is the critical role of GPU compute resources in model fine-tuning and optimization. Leveraging GPUs significantly accelerates hyperparameter experimentation, thereby enhancing efficiency in performance optimization.