A Deep Learning Approach for Skin Lesion Classification on HAM10000 Dataset

Mirza Arslan Baig
AI/ML Engineer
Email: your.email@example.com | GitHub: github.com/yourprofile

Abstract

Early detection and accurate classification of skin lesions, especially melanoma, is critical for patient outcomes. This project focuses on developing a deep learning-based classification model using the HAM10000 dataset. Despite constraints such as limited computational resources (Google Colab free tier) and dataset imbalance, we successfully implemented an EfficientNet-B3 based pipeline achieving around 80% validation accuracy with a macro F1-score of 0.7375. The project also includes visualization tools, model saving, and a Streamlit-ready deployment for demonstration purposes.

1. Introduction

Skin cancer remains one of the most prevalent cancers worldwide. Dermoscopic imaging has improved the diagnostic process, but manual diagnosis remains subjective and prone to errors. Automated skin lesion classification using deep learning has emerged as a promising tool to assist dermatologists. In this project, we address the classification of seven categories of lesions from the HAM10000 dataset using convolutional neural networks with modern transfer learning techniques.

2. Methodology

The workflow was carefully structured to achieve high accuracy under limited GPU resources available on Google Colab. The steps include: 1. **Dataset Preparation**: The HAM10000 dataset was loaded and verified. Images were stored in two directories, requiring preprocessing. 2. **Data Augmentation**: Albumentations library was used with techniques such as random crop, rotation, flips, and color jitter to improve generalization. 3. **Class Imbalance Handling**: WeightedRandomSampler was used to mitigate imbalance among lesion categories. 4. **Model Selection**: EfficientNet-B3 from the `timm` library was chosen for its efficiency and strong baseline performance. 5. **Training Strategy**: Two-phase training was applied: (a) training classifier head, and (b) fine-tuning the entire model. 6. **Optimization**: AdamW optimizer, ReduceLROnPlateau scheduler, and class-weighted cross entropy were used. 7. **Evaluation Metrics**: Accuracy, macro F1-score, precision, recall, and confusion matrix were computed. 8. **Visualization**: Training/validation curves and per-class performance metrics were plotted. 9. **Model Saving**: Final model was exported in Streamlit-ready format for deployment.

3. Experimental Setup

- **Hardware**: Google Colab free tier (Tesla T4 GPU, 16GB RAM). - **Dataset**: HAM10000 with ~10,000 dermatoscopic images across 7 classes. - **Input Size**: 300x300 pixels (EfficientNet-B3 default). - **Batch Size**: 16. - **Epochs**: 6 (head training) + 25 (fine-tuning). - **Software**: PyTorch, Albumentations, timm, scikit-learn, Matplotlib, Seaborn.

4. Results

Our final model achieved nearly **80% validation accuracy** with a **macro F1-score of 0.7375**. Training accuracy reached over 94%, showing the model's ability to learn complex representations. The confusion matrix highlighted the dominance of majority classes such as "bkl", while minority lesion types (nv, mel, bcc, etc.) remained challenging due to dataset imbalance. Training and validation curves showed consistent learning without severe overfitting, validating the effectiveness of augmentation and fine-tuning strategies.

5. Discussion of Challenges

The main difficulties included: - Limited GPU runtime on Colab leading to constrained experimentation. - Severe class imbalance in HAM10000, where certain lesion types had very few samples. - Initial underfitting with simple CNNs, requiring us to adopt EfficientNet transfer learning. - Ensuring reproducibility while handling randomness in augmentation. We overcame these challenges using class balancing, transfer learning, learning rate scheduling, and model checkpointing.

6. Conclusion & Future Work

This project demonstrated that with systematic experimentation and modern deep learning techniques, high accuracy (~80%) can be achieved for multi-class skin lesion classification even under limited computational resources. The trained model was saved in a deployment-ready format and can be integrated into a Streamlit app for interactive demonstration, which makes it suitable for scholarship evaluation. Future work will include: - Extending the pipeline with self-supervised pretraining. - Exploring Vision Transformers (ViT) and ConvNeXt architectures. - Testing on external skin lesion datasets for generalization.