

CSE499A.15

Group 05

Student Name and ID

1. Emon Hossen 2211106042 (used a CLIP-based vision-language pipeline to match herb treatments with skin diseases by encoding herb texts and skin images, computing cosine similarity, ranking the top-5 matches)
2. Faheem Hasnat 2211721642 (implemented the evaluation pipeline, similarity ranking, and results visualization for herb prediction)
3. Kazi Tanora Akther 2132580642 (handled model setup, SigLIP fine-tuning, and feature extraction)

Paragraph 1:(Summary of work)

In this stage, we used a CLIP-based vision-language matching pipeline that filters herb treatments for skin diseases from MeSH classifications, encodes herb descriptions and skin condition images into embeddings, computes cosine similarity scores, and ranks the top-5 most relevant herbal treatments for each skin image with complete visualization and CSV export of results.

Additionally, we trained a SigLIP-based Vision-Language Model (VLM) to connect dermatological image data from Dermnet with textual medicinal herb information from Herb2.0. Using the preprocessed metadata_augmented.csv (Dermnet) and herb2_final_clean.csv (Herb2.0), the model learned cross-modal embeddings that align visual skin disease patterns with herb-related textual descriptions. The purpose of this combined work is to establish an intelligent retrieval framework capable of identifying potential herbal compounds or treatments associated with specific skin diseases.

Paragraph 2:(Future application)

This VLM framework can serve as the foundation for an AI-driven dermatological assistant capable of predicting suitable herbal remedies directly from patient skin images. In future extensions, we plan to integrate SkinCon and HERB ontology data to enhance medical correlations through multimodal fusion and further train the datasets using the BLIP-2 VLM model to improve accuracy and generalization in herb–skin disease prediction.