

DermAssist: An AI-Powered Dermatology Support System

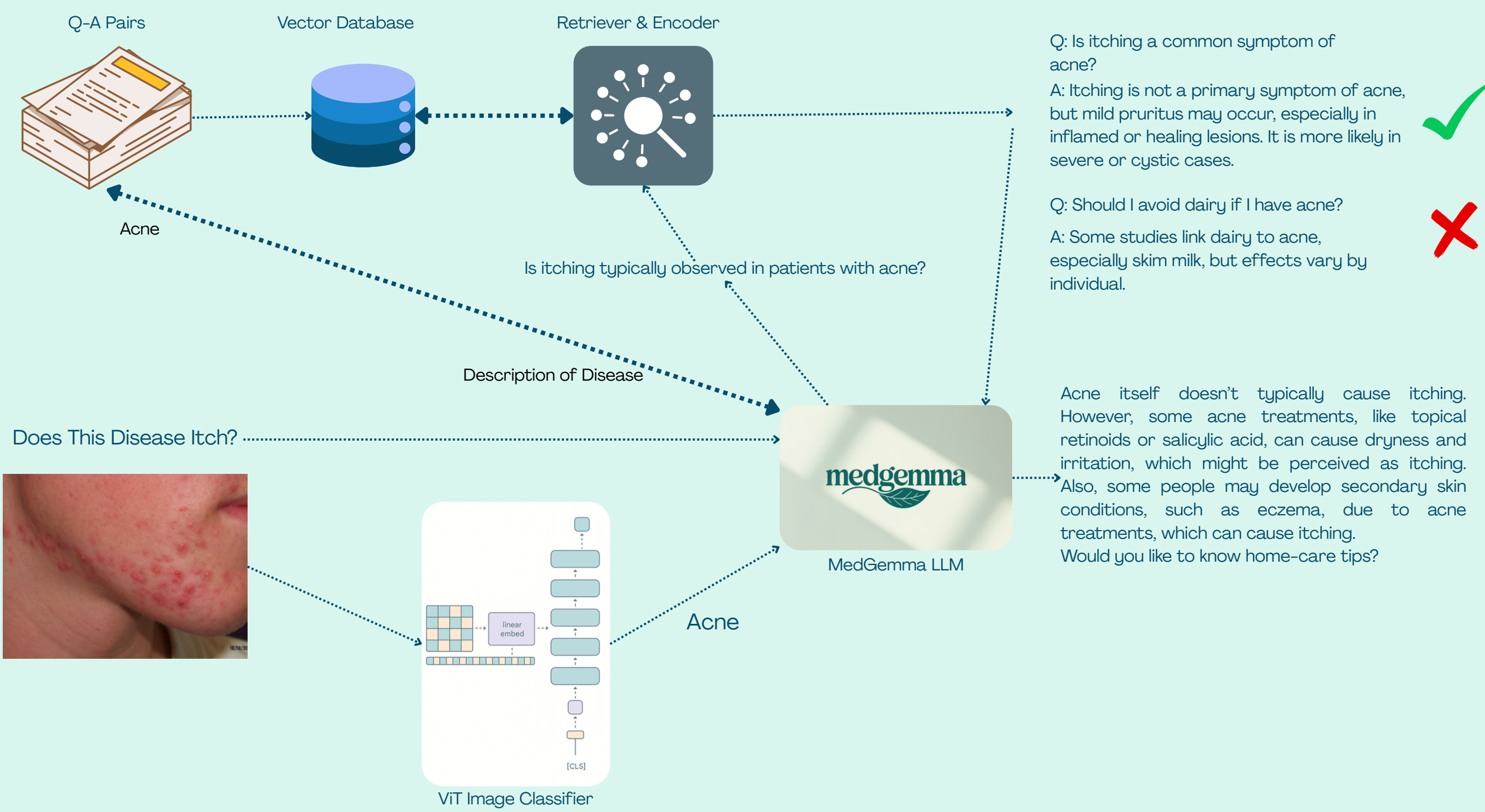
Mehmet Akif Özgür, Bengü Barış Balkan, Furkan Özyurt
Prof. Dr. Nazlı İkizler Cinbiş



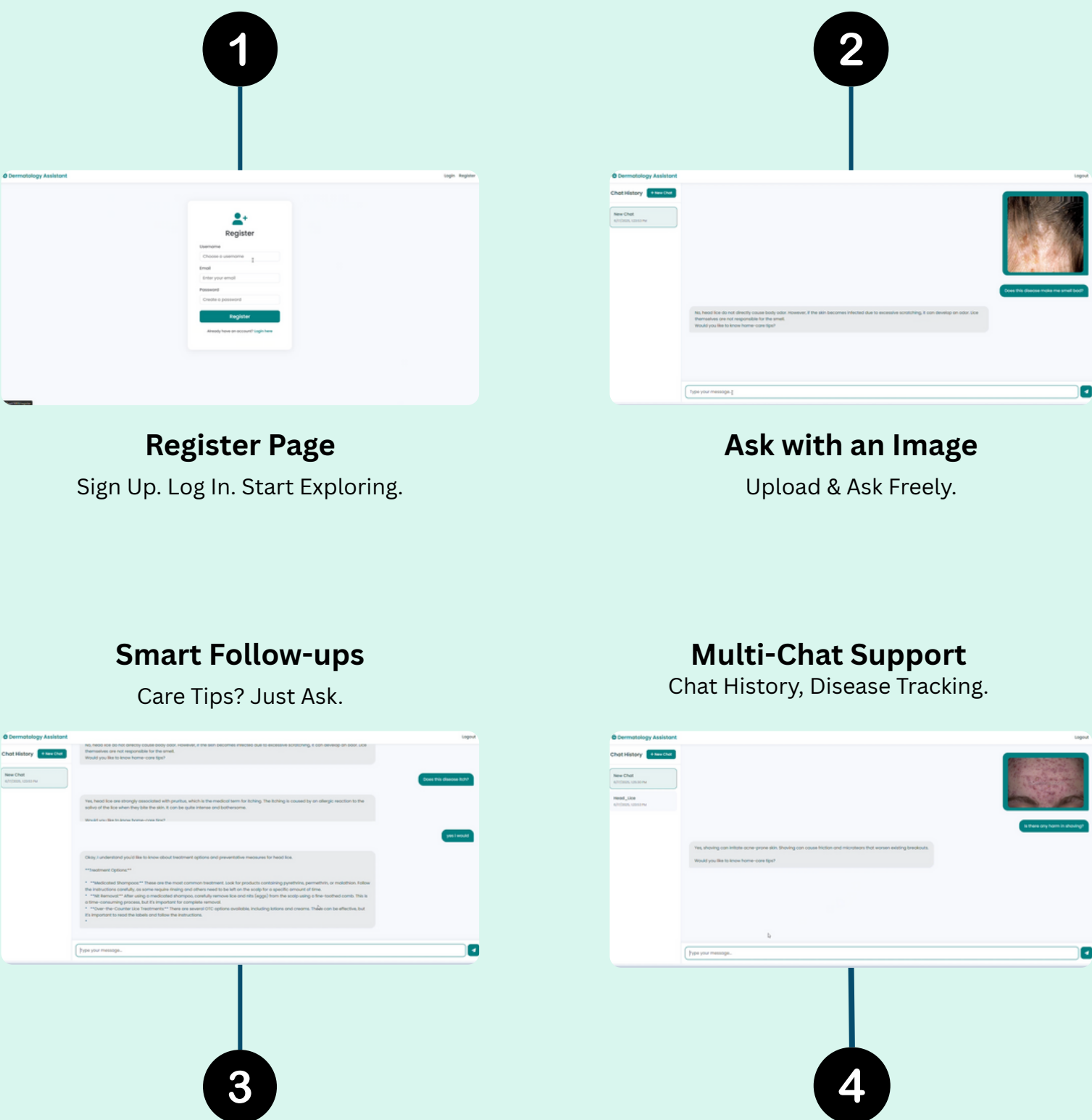
HACETTEPE
UNIVERSITY



End-to-End Processing Pipeline



How the Web Assistant Works



Methodology

Dataset and Image Classifier
We curated a 30-class dermatology dataset and applied inverse class weighting. Images were resized to 224×224 and normalized using ImageNet stats. Among four pretrained models tested—ResNet-50, Swin Transformer, ViT-Base (ImageNet), and ViT-Base (DermNet)—the DermNet-pretrained ViT-Base achieved ~91% validation accuracy and was selected as the final model.

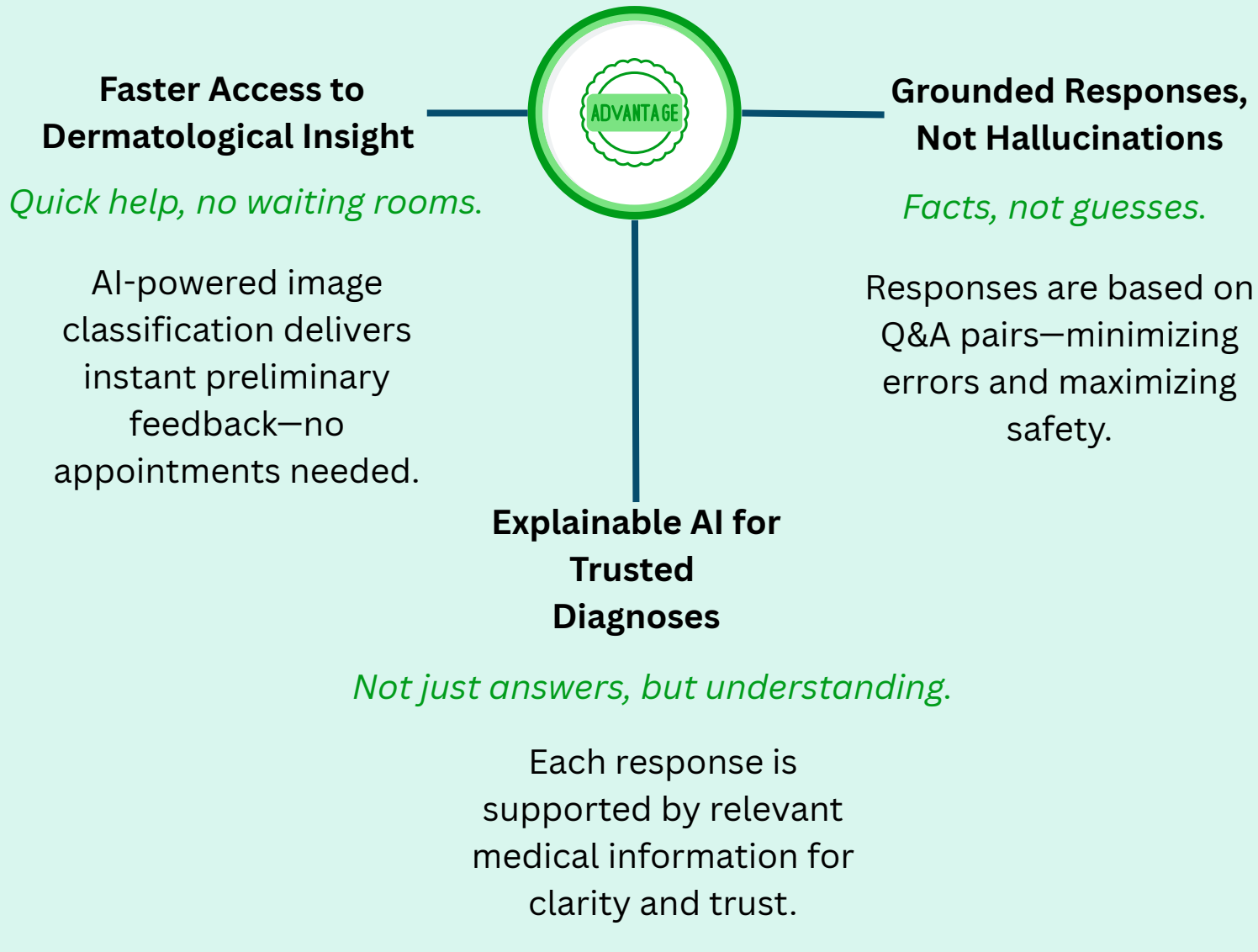
Query Refinement with MedGemma-4b
After classification, the user's question is reformulated by MedGemma-4b into a more medical and precise form—for instance, “Does this disease itch?” becomes “Is itching commonly observed in patients with acne?” This reduces ambiguity for the retrieval step.

Semantic Retrieval
The refined question is embedded using MiniLM-L6-v2 and compared to a database of 50 Q&A pairs per disease class (generated via GPT-3). Cosine similarity determines relevance, and top entries above a threshold are used for context.

Answer Generation
MedGemma-4b then generates a final answer using the original question, its refined version, and the retrieved context. The output is clear, medically accurate, and limited to ~150 tokens unless more detail is needed.

System Deployment
The system runs as a web app. A Flask server handles image uploads and classification, FAISS performs local retrieval, and MedGemma-4b is accessed via API. The chat interface supports real-time, multi-turn conversations.

Key Advantages of Our Assistant



RESULTS

We evaluated the system using the LLM-as-a-judge method on 30 skin conditions, scoring responses for accuracy, empathy, and medical consistency. MedGemma-4b achieved average scores of 8.43 (accuracy), 8.63 (empathy), and 8.43 (consistency). Highest scores were seen in Lupus, Tinea, and Candidiasis, while Vascular and Benign Tumors showed lower performance. Overall, the system delivers reliable, empathetic, and explainable answers.

Traditional NLP metrics like BLEU, ROUGE, and BERTScore yielded low scores, despite accurate responses. These metrics depend on word overlap with reference answers, but our model often uses different yet medically correct phrasing—causing low lexical similarity even when the content is valid.

We used a Vision Transformer (ViT) model initially pre-trained on ImageNet and then fine-tuned on the DermNet dataset. Finally, the model was further fine-tuned using our custom 30-class dermatology dataset. After training, the model achieved an overall accuracy of 91%, with a macro-average F1 score of 87% and a weighted-average F1 score of 91%. These results indicate that the model can classify most skin conditions accurately and consistently. While performance may vary across individual classes, the overall results demonstrate that the model generalizes well across diverse dermatological categories.

MODELS	LLM ACCURACY	LLM EMPATHY	LLM MEDICAL CONSISTENCY	CLASSIFIER ACCURACY	CLASSIFIER MACRO-AVG F1	CLASSIFIER WEIGHTED-AVG F1
MEDGEMMA-4B 8Q	8.43/10	8.63/10	8.43/10	-	-	-
GEMMA-4B 8Q	8.31/10	9.1/10	8.29/10	-	-	-
VIT (DERMNET FINE-TUNED)	-	-	-	91%	87%	91%
BASE VIT	-	-	-	89%	%84	%89

EXAMPLES



Question: Does this disease make me smell bad?
Refined Question: Can head lice cause body odor?
QA Pair:
Q: Do lice cause a bad smell?
A: No, but infected skin from scratching might develop an odor in rare cases.
Answer: No, head lice do not directly cause body odor. However, if the skin becomes infected due to excessive scratching, it can develop an odor. Lice themselves are not responsible for the smell. Would you like to know home-care tips?