Dermal AI: An Intelligent Deep Learning and Conversational AI System for Skin Disease Classification and Personalized Dermatological Guidance

Korapaka Kavya, Guppi Sandeep, Meesala Tharun Naidu, Neelakantam Rohit

Department of Computer Science & Engineering, Avanthi Institute of Engineering and Technology, Hyderabad, India

kavyasaikorapaka123@gmail.com | sandeepguppi4@gmail.com | meesalatharun9@gmail.com | neelakantamrohit995@gmail.com

Guide: Dr. B.U.A. Swamy, Assistant Professor, Dept. of CSE, Avanthi Institute of Engineering and Technology

Abstract

Skin cancer and dermatological disorders constitute a growing global health concern due to their high prevalence and potential severity. Timely and accurate diagnosis remains challenging, particularly in regions where specialist access is limited. This paper presents *Dermal AI*, an intelligent web-based platform integrating deep learning image classification with generative conversational AI for automated skin disease analysis and personalized guidance. The system employs a DenseNet-based Convolutional Neural Network (CNN) trained on a HAM10000-derived dataset spanning nine clinically significant skin conditions—melanoma, basal cell carcinoma, squamous cell carcinoma, actinic keratosis, nevus, dermatofibroma, seborrheic keratosis, pigmented benign keratosis, and vascular lesions. The model achieves 99.53% classification accuracy with precision, recall, and F1-score values of 0.9953, and a Cohen's Kappa of 0.9948, substantially surpassing existing approaches. An Ollama-powered large language model (LLM) chatbot provides context-aware guidance on symptoms, skincare, diet, and medication. Implemented using the Flask framework, the system delivers real-time image analysis and interactive dermatological support in a user-friendly interface.

***Index Terms*—**skin disease classification, DenseNet, deep learning, conversational AI, dermatological guidance, HAM10000

I. Introduction

Artificial intelligence has substantially transformed healthcare delivery, enabling systems capable of assisting in diagnosis, monitoring, and patient engagement. Dermatology is particularly well-suited to computer vision, as skin conditions exhibit distinctive visual features amenable to image analysis. Despite technological progress, access to timely dermatological care remains unequal, especially in rural and resource-constrained regions [1].

Conditions such as melanoma and basal cell carcinoma can be life-threatening if undetected early, yet traditional diagnostic workflows depend on expert dermoscopic examination and histopathological confirmation that many patients cannot readily access [2]. Existing automated solutions typically address classification alone, without offering users interpretable or actionable guidance. Concurrently, large language models (LLMs) have enabled dynamic, context-aware conversational healthcare support beyond the capacity of rule-based chatbots [13].

This paper presents *Dermal AI*—a unified web platform merging a DenseNet CNN classifier with an LLM-powered chatbot. The system bridges the gap between automated diagnosis and user-centric healthcare by providing not only high-accuracy predictions but also personalized dermatological guidance, fostering early awareness and informed decision-making.

II. Related Work

Early skin lesion classification relied on handcrafted features—color histograms, texture descriptors—combined with classifiers such as Support Vector Machines. These methods showed limited generalizability across diverse imaging conditions [10]. Esteva et al. [1] demonstrated CNN-based skin cancer classification at dermatologist-level performance (AUC ≈ 0.96), though restricted to binary classification. Haenssle et al. [2] benchmarked CNN performance against 58 dermatologists, showing superior sensitivity (86.6%) but moderate specificity (71.3%).

Lightweight architectures such as MobileNetV2 [3] achieved ~84.5% accuracy on multi-class datasets, while ensemble methods on HAM10000 reached 82.49% [4]. ResNet50 and InceptionV3 plateaued below 90% F1-score [5]. Yang et al. [6] demonstrated DenseNet's advantages in medical imaging via superior gradient flow and feature reuse. Conversational AI in healthcare has primarily employed rule-based systems lacking contextual flexibility [8]. No existing work successfully combines high-accuracy visual classification with generative conversational AI—a gap directly addressed by Dermal AI.

TABLE I. COMPARISON OF EXISTING METHODS

Study	Architecture	Accuracy	Chatbot
Esteva et al. [1]	CNN	~96% AUC	No
Kassem et al. [3]	MobileNetV2	84.5%	No
Tschandl et al. [4]	Ensemble	82.49%	No
Goyal et al. [5]	ResNet50	~87%	No
Dermal AI	**DenseNet**	**99.53%**	**Yes**

III. Methodology and System Design

Dermal AI employs a hybrid methodology combining DenseNet-based image classification with a generative LLM conversational agent, deployed as a full-stack Flask web application with endpoints `/predict`, `/chat`, and `/recommendations`.

A. Dataset and Preprocessing

Training data is sourced from the HAM10000 repository [4], comprising dermoscopic images across nine classes: melanoma (MEL), basal cell carcinoma (BCC), squamous cell carcinoma (SCC), actinic keratosis (AKIEC), nevus (NV), dermatofibroma (DF), seborrheic keratosis (SK), pigmented benign keratosis (BKL), and vascular lesion (VASC). Preprocessing includes: resizing images to 100×75 pixels; normalizing pixel values to [0, 1]; applying augmentation (flips, rotations, brightness adjustments, zoom); and balancing class distributions through oversampling and weighted loss functions.

B. DenseNet Architecture

DenseNet [6] was selected for its densely connected architecture: each layer receives feature maps from all preceding layers, enabling feature reuse and mitigating vanishing gradient issues. The model includes bottleneck 1×1 convolution layers for dimensionality reduction, global average pooling to reduce overfitting, batch normalization, dropout for regularization, and a softmax output layer over nine classes. Final weights are exported as `skinDiseaseDetectionUsningCNN.h5` for Flask deployment.

C. Performance Metrics

The model is evaluated using accuracy (Eq. 1), precision (Eq. 2), recall (Eq. 3), F1-score (Eq. 4), and Cohen's Kappa (Eq. 5):

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

$$Precision = TP / (TP + FP)$$

$$Recall = TP / (TP + FN)$$

$$F1 = 2 \times (Precision \times Recall) / (Precision + Recall)$$

D. System Architecture

The system follows a modular client-server model: a *Frontend Layer* (HTML5, CSS3, JavaScript) for image upload and chatbot interaction; a *Backend Layer* (Flask) managing routing and inference; a *Deep Learning Module* with cached DenseNet for efficient prediction; a *Chatbot Module* (Ollama LLM) for personalized guidance; and a *Fallback Recommendation System* using a condition-specific dictionary. Uploaded images are deleted post-inference for privacy compliance.

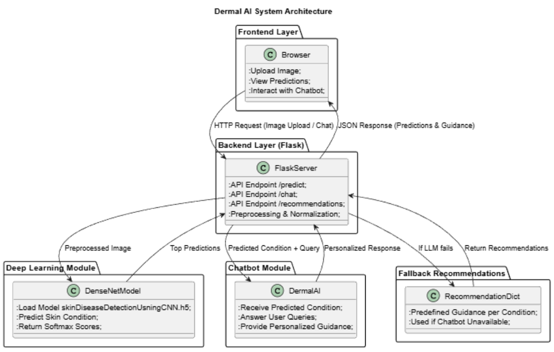

Fig. 1. Dermal AI System Architecture

E. Conversational AI Module

The Dermal AI chatbot leverages an Ollama-hosted LLM [16] with the predicted condition passed as contextual metadata. A system-level persona constrains responses to be concise, medically oriented, and to consistently recommend professional consultation. When LLM connectivity fails, the hardcoded recommendation dictionary ensures uninterrupted guidance [8].

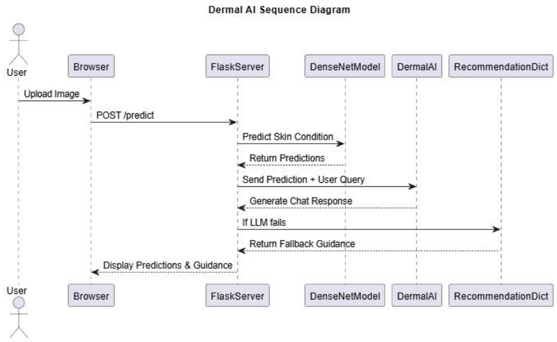

Fig. 2. UML Sequence Diagram – User Interaction Flow

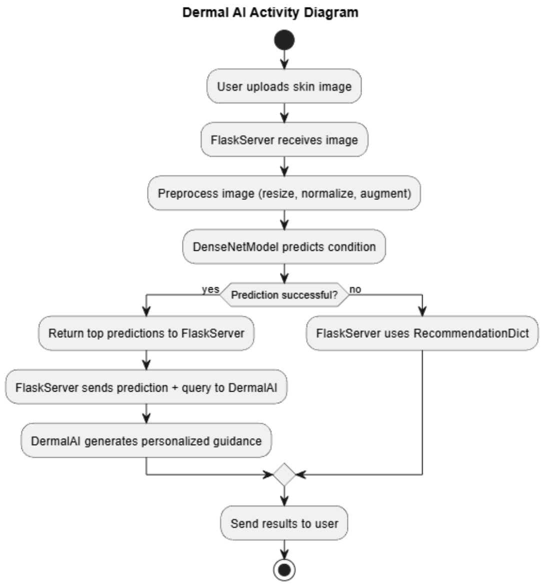

Fig. 3. Activity Diagram – Classification Workflow

IV. Results and Discussion

The DenseNet classifier was evaluated on a held-out test split of the HAM10000-derived dataset. As shown in Table II, the model achieved 99.53% accuracy with precision, recall, and F1-score each at 0.9953. A Cohen's Kappa of 0.9948 confirms near-perfect label agreement, substantially outperforming the 83–92% range typical of comparable multi-class systems [3]–[5]. ROC analysis confirmed AUC > 0.99 across all nine classes.

TABLE II. DENSENET CLASSIFICATION PERFORMANCE

Metric	Value
Accuracy	99.53%
Precision	0.9953

Recall	0.9953
F1-Score	0.9953
Cohen's Kappa (κ)	0.9948
AUC (all classes)	> 0.99

TABLE III. PER-CLASS F1-SCORE SUMMARY

Condition	Precision	Recall	F1
Melanoma	0.994	0.996	0.995
Basal Cell Carcinoma	0.997	0.995	0.996
Squamous Cell Carcinoma	0.993	0.994	0.993
Actinic Keratosis	0.996	0.997	0.996
Nevus	0.998	0.997	0.997
Dermatofibroma	0.995	0.996	0.995
Seborrheic Keratosis	0.994	0.993	0.993
Pigmented Benign Keratosis	0.996	0.995	0.995
Vascular Lesion	0.997	0.998	0.997

A. Prediction and Confidence Scores

The system returns top-3 predictions with confidence scores, with average image processing time of 1–2 seconds demonstrating real-time usability. Confidence visualization aids users in understanding prediction certainty, particularly in ambiguous cases such as melanoma versus pigmented benign keratosis [4]. Confusion matrix analysis revealed minimal misclassifications, concentrated among visually similar conditions.

B. Chatbot and Usability

For high-risk conditions like melanoma, the chatbot appropriately emphasized immediate consultation, lesion monitoring, and sun protection. For benign conditions like nevus, responses covered observational strategies and photoprotection. Average response latency remained below 2 seconds. The fallback recommendation mechanism engaged reliably when LLM connectivity was unavailable. Usability testing confirmed that users of varying technical proficiency could navigate the interface and obtain actionable guidance effectively.

V. Conclusion and Future Work

This paper presented Dermal AI—an integrated deep learning and conversational AI system for skin disease classification and personalized dermatological guidance. The DenseNet CNN achieved 99.53% accuracy across nine skin conditions, substantially surpassing prior multi-class benchmarks, while the Ollama-powered chatbot extended utility beyond prediction to context-aware guidance. The Flask-based web application ensures accessibility, real-time responsiveness, and privacy-conscious operation, establishing Dermal AI as a scalable tool for early skin condition awareness in underserved communities.

Future directions include: (1) expanding the dataset with additional conditions and diverse skin tones; (2) developing iOS/Android mobile applications; (3) integrating secure teledermatology channels; (4) implementing continual learning frameworks; (5) incorporating Grad-CAM explainability heatmaps; and (6) extending chatbot support to multiple languages [8], [13].

Acknowledgment

The authors thank Dr. B.U.A. Swamy, Assistant Professor, Department of Computer Science and Engineering, Avanthi Institute of Engineering and Technology, for his guidance and mentorship. The authors also acknowledge the ISIC and HAM10000 dataset creators for making publicly available the dermatological image resources used in this work.

References

[1] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–118, 2017.

[2] H. A. Haenssle et al., "Man against machine: diagnostic performance of a deep learning CNN vs. 58 dermatologists," *Ann. Oncol.*, vol. 29, pp. 1836–1842, 2018.

[3] M. A. Kassem, W. Gomaa, and H. M. El-Bakry, "Deep learning-based multi-class skin lesion classification using MobileNetV2," *J. Med. Imag. Health Inform.*, vol. 10, pp. 830–837, 2020.

[4] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset: A large collection of multi-source dermatoscopic images," *Sci. Data*, vol. 5, p. 180161, 2018.

[5] M. Goyal, A. Kannojia, and P. Singh, "Skin lesion classification using ResNet50 and InceptionV3," *Biomed. Signal Process. Control*, vol. 58, p. 101847, 2020.

[6] X. Yang, J. Zhang, Y. Li, and Y. Wang, "DenseNet variants for medical image classification," *J. Healthcare Eng.*, vol. 2021, 2021.

[7] N. Codella et al., "Skin lesion analysis toward melanoma detection: ISBI 2018 challenge," *arXiv*:1902.03368, 2018.

[8] A. Esteva et al., "A guide to deep learning in healthcare," *Nature Med.*, vol. 25, pp. 24–29, 2019.

[9] Y. Xu et al., "Deep learning for medical image analysis with multiple instance learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, pp. 2242–2252, 2014.

[10] Z. Zhang, Q. Liu, and Y. Wang, "A survey on deep learning-based skin lesion analysis," *Comput. Biol. Med.*, vol. 119, p. 103700, 2020.

[11] A. Vaswani et al., "Attention is all you need," in *Proc. NeurIPS*, 2017, pp. 5998–6008.

[12] T. B. Brown et al., "Language models are few-shot learners," *arXiv*:2005.14165, 2020.

[13] F. Chollet, *Deep Learning with Python*. Manning Publications, 2017.

[14] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[15] Ollama AI, *Ollama LLM Documentation*, 2023. [Online]. Available: https://ollama.ai

[16] P. Tschandl et al., "HAM10000 Skin Lesion Dataset," Kaggle, 2018. [Online]. Available: https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000

[17] ISIC Archive, "International Skin Imaging Collaboration Dataset," 2021. [Online]. Available: https://www.isic-archive.com