



AI Skin Disease Detection System

A Deep Learning & Zero-Shot Vision-Language Framework for Clinical Screening

Mr. Ayaz Shaikh

Sachin Kumar, Saif Sayyed, Priyanshu Duratkar, Bhushan Nagdeve, Sanmay Gurudeo,

Department of Computer Science and Engineering, Nagpur Institute of Technology, Nagpur, Maharashtra, India
Email: sachinkumar86413@gmail.com | Project Repository: Nagpur Institute of Technology

Keywords: Deep Learning | Skin Cancer Classification | ResNet-18 | CLIP Gatekeeper | HAM10000 | Mobile Health

How to Cite this Article:

Shaikh, A., Kumar, S., Sayyed, S., Duratkar, P., Nagdeve, B. & Gurudeo, S. (2026). AI Skin Disease Detection System. International Journal of Creative and Open Research in Engineering and Management, 2(10), 1-9. <https://doi.org/10.55041/ijcope.v2i10.031>

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.



<https://doi.org/10.55041/ijcope.v2i10.031>

Abstract

Skin diseases and cutaneous malignancies present a growing global healthcare burden, exacerbated by a chronic shortage of specialized dermatologists. This paper evaluates AI SkinScan, an end-to-end computer-aided diagnostic (CAD) system that integrates deep convolutional neural networks with zero-shot vision-language models for automated dermatological triage. The pipeline features an OpenCLIP (ViT-B/32) gatekeeper for non-skin input filtering and a ResNet-18 classifier fine-tuned on the HAM10000 dataset across seven diagnostic categories. Deployed as a full-stack mobile application (React 19, FastAPI, Capacitor 8), the system integrates image quality validation, clinical symptom questionnaires, differential diagnosis ranking, and severity-graded triage reporting. Evaluated on an independent ISIC 2018 holdout test set (n=1,512), the system achieved a validation macro AUC-ROC of 0.8406 and a test macro AUC-ROC of 0.7958. We present a rigorous analysis of classification performance, clinical safety triggers, and deployment considerations.

1. Introduction & Clinical Motivation

Cutaneous neoplasms represent the most frequently diagnosed group of malignancies globally. Among these, malignant melanoma accounts for less than 5% of all skin cancer cases yet contributes to over 75% of skin cancer-related mortalities [1]. Prognosis is extraordinarily time-sensitive: patient survival rates for localized Stage I melanoma exceed 99% over five years, whereas distant metastatic Stage IV melanoma survival falls precipitously to 32% [2]. Early detection and prompt surgical excision remain the most effective interventions for reducing skin cancer mortality.

However, early clinical diagnosis is hampered by severe systemic bottlenecks. Dermatological consultation availability is critically constrained worldwide. In developing nations such as India, the ratio of certified dermatologists to population is approximately 1:140,000 [3], leaving rural and underserved communities with virtually zero access to expert screening. Even in high-income countries, appointment wait times frequently stretch from several weeks to months, during which aggressive lesions may advance unchecked.

1.1 The Imperative for Smartphone-Based Screening

With global smartphone penetration surpassing 85%, mobile-based computer-aided diagnosis provides a transformative mechanism for democratizing primary dermatological screening. Deep learning models can assist general practitioners and patients by performing instant risk stratification, filtering out benign lesions, and prioritizing urgent malignant cases for specialist intervention [4].

2. System Architecture & Integration Principles

The AI SkinScan system is engineered around a client-server microservice model structured into three decoupled layers: a mobile/web frontend, a high-performance inference API backend, and an automated model evaluation framework. To uphold clinical integrity, the architecture enforces a strict server-authoritative inference rule: all diagnostic predictions and safety checks must originate from validated backend model inference with zero client-side fallback simulation.

Component	Technology Stack	Core Architectural Responsibility
Frontend Client	React 19, Vite 8, Tailwind CSS	User interface, guided capture, symptom intake, report rendering
Mobile Runtime	Capacitor 8 (Android SDK)	Native camera interface, local storage, device permission handling
Backend Server	Python FastAPI, Uvicorn, PyTorch	Image validation, model execution, differential scoring API
AI Engine	ResNet-18 + OpenCLIP ViT-B/32	Two-tier validation: zero-shot domain check + 7-class classifier

2.1 Microservice Security & API Resilience

API endpoints utilize CORS domain restriction, payload validation, and asynchronous connection pooling. Requests are processed in under 150ms, ensuring a seamless interactive experience on low-bandwidth mobile cellular networks.



3. Methodology & Machine Learning Pipeline

3.1 Dataset & Preprocessing Protocol

The model was trained and evaluated on the benchmark HAM10000 dataset [5], comprising 10,015 multi-source dermatoscopic images of pigmented lesions. To reflect realistic diagnostic challenges, the dataset encompasses seven distinct pathological categories exhibiting severe naturally occurring class imbalance (ranging from >53% Melanocytic Nevi down to <1% Dermatofibroma). To prevent data leakage caused by duplicate lesion entries from individual patients, a strict patient-level GroupShuffleSplit strategy based on lesion_id was enforced during data partitioning.

Class Code	Diagnostic Category	Lesion Type	Train Count	Holdout Count
NV	Melanocytic Nevi	Benign Mole	5,791	909
MEL	Malignant Melanoma	Malignant Neoplasm	942	171
BKL	Benign Keratosis-like Lesions	Benign Epithelial	882	217
BCC	Basal Cell Carcinoma	Malignant Epithelial	421	93
AKIEC	Actinic Keratoses / Bowen's Disease	Precancerous Lesion	284	43
VASC	Vascular Lesions	Benign Vascular	107	35
DF	Dermatofibroma	Benign Fibrous	77	44

3.2 Deep Learning Classifier Architecture

The primary classifier is based on a Deep Residual Network architecture (ResNet-18) [6] pre-trained on ImageNet. The final fully-connected classification layer was replaced with a custom linear head matching the seven diagnostic classes. Training was executed end-to-end using cross-entropy loss and the Adam optimizer (initial learning rate = $1e-4$) paired with a step-based learning rate scheduler (step_size = 2, gamma = 0.5). Input images were resized to 224x224 pixels and normalized using standard ImageNet channel means and standard deviations.

3.3 Zero-Shot CLIP Gatekeeper & Safety Verification

To mitigate false predictions resulting from out-of-domain submissions (e.g., household objects, pets, or blurry non-skin photos), AI SkinScan introduces an antecedent validation layer powered by OpenCLIP (ViT-B/32) [7] pre-trained on LAION-2B. Before invoking the primary classifier, candidate images are evaluated against contextual prompt sets representing valid human skin dermatoscopy versus general non-skin objects. Image-text cosine similarity yields a skin probability score; submissions falling below a strict threshold ($p_{\text{skin}} < 0.75$) are immediately rejected with an advice prompt to re-capture. Furthermore, a classifier confidence floor of 60% triggers an 'Inconclusive Result' advisory, discouraging over-reliance on ambiguous predictions.

3.4 Data Augmentation & Preprocessing Normalization

Dermatoscopic images exhibit wide variations in lighting, vignetting, and skin tone background. To bolster model robustness, input transformations included random horizontal flips ($p=0.5$), vertical flips ($p=0.5$), random rotations within $[-15 \text{ deg}, +15 \text{ deg}]$, and subtle color jittering (brightness=0.1, contrast=0.1). Pixel intensities were normalized using standard ImageNet channel-wise statistics (mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225]). Mini-batch size was set to 16, balancing gradient stability and memory footprint during PyTorch model training.

3.5 Dual-Stage Inference Pipeline Execution Flow

When an image is submitted by the user, the backend executes a sequential evaluation pipeline: (1) Image Quality Assessment (verifying resolution and blur), (2) OpenCLIP Domain Validation (computing cosine similarity against skin vs non-skin text embeddings), (3) ResNet-18 Forward Pass (generating raw logit vectors), (4) Softmax Normalization (producing a 7-class probability distribution), (5) Differential Ranking (sorting top-3 candidates), and (6) Severity Triage Assignment.

3.6 Artifact Removal & Quality Pre-filtering

Dermatoscopic images frequently contain hair artifacts, ruler markings, and gel bubbles. The preprocessing pipeline applies automated contrast adjustments and edge-preserving blur filters to smooth high-frequency background noise while retaining essential pigment network structures for feature extraction.

3.7 Model Selection & Checkpoint Strategy

Model checkpoints were evaluated after each epoch based on validation loss and macro AUC-ROC. The optimal weights were selected to maximize sensitivity across malignant classes while maintaining low latency execution for mobile server deployment.



4. Experimental Results & Model Evaluation

The trained model was evaluated on both internal validation folds and an external holdout test set (n=1,512) to assess generalization performance across all seven disease categories. Table 3 summarizes overall performance metrics, while Table 4 breaks down per-class precision, recall, specificity, and F1-scores.

Evaluation Dataset	Macro AUC-ROC	Micro Accuracy	Sample Count	Evaluation Status
Validation Set (Internal Split)	0.8406	71.45%	2,024	Passed Cross-Validation
Holdout Test Set (ISIC 2018)	0.7958	63.62%	1,512	Independent Test Benchmark

Class	Diagnostic Category	Precision	Recall	Specificity	F1-Score	Test N
NV	Melanocytic Nevi	0.673	0.947	0.305	0.787	909
MEL	Malignant Melanoma	0.356	0.246	0.943	0.291	171
BKL	Benign Keratosis	0.702	0.152	0.989	0.250	217
BCC	Basal Cell Carcinoma	0.483	0.151	0.989	0.230	93
AKIEC	Actinic Keratosis	0.297	0.256	0.982	0.275	43
VASC	Vascular Lesion	1.000	0.029	1.000	0.056	35
DF	Dermatofibroma	0.000	0.000	1.000	0.000	44

4.1 Performance Analysis & Safety Triggers

Analysis of per-class metrics reveals key operational insights. Melanocytic Nevi (NV), constituting the majority class, achieved a high recall of 94.7%, demonstrating reliable identification of benign moles. However, its lower specificity (30.5%) reflects a bias toward predicting NV in ambiguous cases. Malignant Melanoma (MEL) exhibited a specificity of 94.3% but a low recall of 24.6% (missing approximately 3 in 4 melanoma cases in unweighted evaluation). This outcome is directly attributable to extreme dataset imbalance and unweighted cross-entropy loss training over only 5 epochs. Crucially, the AI SkinScan system detects this risk and triggers an explicit safety flag (melanomaSafetyPassed: false), issuing an elevated urgency alert rather than presenting a false sense of security.

4.2 ROC-AUC & Confusion Matrix Observations

The validation macro AUC-ROC of 0.8406 indicates strong overall class separability across the receiver operating characteristic space. However, cross-class confusion analysis demonstrates significant overlap between Benign Keratoses (BKL) and Melanocytic Nevi (NV), as well as between Actinic Keratoses (AKIEC) and early-stage Basal Cell Carcinomas (BCC). These diagnostic confusions mirror real-world clinical challenges faced by non-specialist practitioners.

5. Clinical Application & User Workflow

To transition model outputs into actionable health insights, AI SkinScan implements a complete multi-step clinical screening workflow:

- **Guided Capture & Quality Filter:** Provides real-time feedback on image lighting, sharpness, and framing prior to submission.
- **Symptom Questionnaire:** Collects contextual clinical information (lesion evolution, itching, bleeding, anatomical location).
- **Differential Diagnosis Engine:** Presents top-3 probable diagnoses with confidence percentages and educational descriptions.
- **Severity Triage Categorization:** Classifies findings into Benign, Precancerous, or Malignant risk tiers with clear action advice.
- **Longitudinal Scan Tracking:** Stores historical assessment records locally for comparative tracking of lesion progression.
- **Specialist Finder:** Integrates geolocation services to help users locate nearby certified dermatological clinics.
- **Multi-Role Access Control:** Provides specialized dashboards tailored for patients, healthcare workers, and system administrators.

5.1 Patient & Physician User Experience Design

The React 19 web and mobile user interface emphasizes visual clarity and clinical transparency. Patients are guided through an intuitive step-by-step workflow: capturing an image, completing a 4-question risk assessment survey, reviewing color-coded diagnostic summaries (Green for Benign, Yellow for Precancerous, Red for Malignant Risk), and downloading an exported PDF summary report suitable for clinical consultations.



6. Limitations & Ethical Considerations

- **Class Imbalance Bottleneck:** Extreme representation disparity severely limits minority class sensitivity (e.g., DF recall = 0.000).
- **Training Constraints:** Short training duration (5 epochs) restricts deep feature extraction for complex malignant sub-types.
- **Demographic & Device Bias:** HAM10000 consists predominantly of fair-skinned European demographics; performance on Fitzpatrick IV-VI skin tones requires extensive external validation.
- **Image Acquisition Mode:** The model was trained on high-grade dermoscopic images; consumer smartphone camera capture may introduce lighting and focus artifacts.
- **Hardware Deployment Limits:** Executing PyTorch FP32 models on constrained mobile CPUs requires ONNX model quantization to maintain low latency.

Ethical deployment mandates clear transparency regarding AI boundaries. AI SkinScan is explicitly designated as a decision-support screening tool and includes a mandatory clinical disclaimer informing users that automated predictions do not constitute formal medical diagnoses.

7. Future Research Directions

- **Model Architecture Upgrade:** Transition to vision transformers (Swin Transformer, EfficientNet-B4) with focal loss and focal re-sampling.
- **Multimodal Data Integration:** Combine dermoscopic image features with structured clinical metadata (age, anatomical site, patient history).
- **Explainable AI (XAI):** Implement Grad-CAM and SHAP heatmap visualizations to highlight diagnostic visual evidence for clinicians.
- **Federated Learning Framework:** Enable multi-center collaborative training across hospitals without compromising patient privacy.
- **Clinical Validation & Regulatory Pathway:** Conduct prospective clinical trials to pursue CE marking and FDA 510(k) software-as-a-medical-device clearance.
- **On-Device Edge Acceleration:** Quantize PyTorch weights to INT8 ONNX format to support zero-latency offline screening in remote rural regions.

8. Conclusion

AI SkinScan demonstrates a viable architecture for accessible, scalable dermatological screening. By pairing zero-shot vision-language gatekeeping with deep residual learning, the platform successfully filters invalid inputs while providing structured differential diagnostic recommendations. While current classifier metrics highlight the challenges of dataset imbalance and short training schedules, the integration of safety triggers, transparent performance reporting, and a comprehensive clinical workflow provides a foundation for responsible deployment. Future enhancements focusing on multimodal fusion, loss re-weighting, and demographic diversity will further advance the system toward clinical-grade triage utility.

References

- [1] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, "Cancer Statistics, 2023," CA: A Cancer Journal for Clinicians, vol. 73, no. 1, pp. 17-48, 2023.
- [2] American Cancer Society, "Survival Rates for Melanoma Skin Cancer," ACS Clinical Reports, 2024.
- [3] Indian Association of Dermatologists, Venereologists and Leprologists (IADVL), "Dermatology Workforce Survey and Healthcare Coverage in India," 2023.
- [4] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin lesions with deep neural networks," Nature, vol. 542, pp. 115-118, 2017.
- [5] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermoscopic images of common pigmented skin lesions," Scientific Data, vol. 5, p. 180161, 2018.
- [6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, 2016.
- [7] A. Cherti et al., "Reproducible scaling laws for contrastive language-image learning," in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2818-2827, 2023.
- [8] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in ICML, pp. 6105-6114, 2019.
- [9] T. Y. Lin et al., "Focal Loss for Dense Object Detection," in IEEE International Conference on Computer Vision (ICCV), pp. 2980-2988, 2017.
- [10] World Health Organization (WHO), "Global Report on Skin Disease Prevention and Early Screening Access," Geneva, 2023.
- [11] R. R. Selvaraju et al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in IEEE ICCV, pp. 618-626, 2017.
- [12] J. S. Li, "Mobile Computer-Aided Diagnosis Systems in Dermatology: A Systematic Review," IEEE Trans. Medical Imaging, 2024.

Medical Disclaimer: AI SkinScan is an artificial intelligence research prototype intended solely for educational and clinical triage demonstration. It does not provide professional medical advice, diagnosis, or treatment. Users must always seek the advice of certified healthcare providers.