

DermoAI

AI-Powered Skin Disease Triage System

Build the Future with AI 2026

Prepared by

Ishan Asati
Aarush Sonkar

Project Resources

GitHub Repository: <https://github.com/IshanAsati/dermoai-final/>

Table of Contents

1 Executive Summary	2
2 Problem Statement	3
2.1 Background	3
2.2 Limitations of Existing Solutions	3
2.3 Objectives	3
3 Methodology	3
3.1 Phase 1: Baseline Model	3
3.2 Phase 2: Architecture Upgrade	3
3.3 Phase 3: Data Leakage Remediation	3
3.4 Phase 4: Domain Gap Analysis	3
3.5 Phase 5: Domain Adaptation	4
4 System Architecture	4
5 User Interface & Demo	5
6 Dataset Engineering	5
6.1 Dataset Composition	5
6.2 Data Quality Measures	6
7 Model Architecture	6
8 Results	6
8.1 Overall Performance	6
9 Cancer Detection Analysis	7
10 Error Analysis	7
10.1 Primary Failure Mode	7
10.2 Secondary Confusions	8
11 Comparative Advantage	8
12 Limitations	8
13 Future Work	8
14 Impact	9
14.1 Target Users	9
14.2 Anticipated Benefits	9
15 Conclusion	9
16 Project Resources	10

1 Executive Summary

DermoAI is a clinical-grade artificial intelligence platform developed to address global disparities in dermatological access by screening for skin diseases and triaging high-risk malignancies from consumer smartphone photographs. Designed as a safety-first clinical decision support and triage system, DermoAI does not replace clinical diagnosis but rather acts as an accessible gateway, flagging high-risk cases to accelerate specialist referrals and optimize clinical pathways.

Leveraging a robust deep learning backbone (ConvNeXt-Base) trained on a curated, multi-source hybrid dataset of over 20,000 dermatological images, the model has been rigorously validated on a clean, deduplicated set of 4,555 validation images spanning 12 distinct skin disease categories. DermoAI demonstrates a multi-class classification accuracy of 82.79%, a binary triage sensitivity of 87.30% for high-risk malignancies, and a specificity of 92.27%.

82.79%

Overall Accuracy

87.30%

Cancer Sensitivity

92.27%

Cancer Specificity

12

Disease Classes

This report documents the system architecture, dataset engineering methodology, model development lifecycle, performance evaluation, and clinical implications of the DermoAI platform.

2 Problem Statement: The Healthcare Access Gap

2.1 Background

Skin cancer remains one of the most prevalent yet preventable forms of malignancy globally. However, access to specialized dermatological care is severely constrained. Patients in rural, underserved, or socioeconomically disadvantaged regions face substantial barriers to care, including prohibitive wait times exceeding several months, high clinical consultation costs, and geographic isolation. This diagnostic delay directly correlates with poorer clinical

outcomes, particularly for aggressive malignancies like Melanoma, where early detection is paramount.

2.2 Limitations of Existing Solutions

Current artificial intelligence solutions in mobile dermatology remain under-utilized in clinical pathways due to several technical bottlenecks:

- **Narrow Disease Coverage:** Most models are restricted to binary classification (e.g., melanoma vs. nevus) rather than the broad differential diagnoses encountered in primary care.
- **Dataset Homogeneity:** Models trained on uniform clinical populations fail to generalize to diverse skin tones and real-world imaging environments.
- **Clinical-to-Smartphone Domain Shift:** A severe drop in accuracy occurs when models evaluated on clinical, dermoscopic photography are exposed to noisy, smartphone-captured images.
- **Lack of Triage Safety:** Systems often attempt to replace diagnostic clinicians rather than serve as a conservative screening filter, raising regulatory and safety concerns.
- **Lack of Explainability:** Black-box predictions do not provide visual feedback, limiting dermatologist trust.
 - Limited disease coverage, often restricted to a small number of categories
 - Training on small or homogeneous datasets that do not generalise to real-world conditions
 - Poor performance on smartphone-captured images due to domain shift
 - Lack of clinical triage capability to differentiate between benign and malignant conditions
 - Limited transparency and explainability in model decision-making

2.3 Objectives

This project aims to develop DermoAI, a production-ready, safety-first skin disease triage platform designed to:

- Map smartphone-captured lesion photographs to a robust 12-class differential diagnosis.
- Optimize triage sensitivity for malignant and pre-malignant lesions to minimize false negatives.
- Bridge the clinical-to-smartphone domain gap through targeted data engineering and domain adaptation.
- Generate structured, clinical reports to guide patients safely into appropriate medical channels.
 - Achieves clinically meaningful accuracy across a broad range of skin conditions
 - Maintains high sensitivity for malignant lesions to minimise false negatives
 - Generalises to real-world smartphone imagery
 - Provides actionable recommendations to guide users toward appropriate care

3 Methodology: The Rigorous Development Journey

The development of DermoAI followed an empirical, five-phase R&D cycle. Each phase was systematically designed to uncover and remediate core failure modes identified during validation, shifting the system from a fragile academic model to a robust, deployment-ready triage platform.

3.1 Phase 1: Baseline Model

Phase 1: Baseline Model (ResNet-50)

An initial baseline classifier was implemented using a pretrained ResNet-50 architecture. While the model achieved a moderate overall validation accuracy of 68.00%, error analysis revealed a critical failure to capture fine-grained textural patterns and micro-vascular structures essential for differentiating malignant lesions from benign lookalikes.

3.2 Phase 2: Architecture Upgrade

Phase 2: Architecture Upgrade (ConvNeXt-Base)

To resolve the limitations of the baseline, the backbone was upgraded to ConvNeXt-Base (88M parameters). This modern convolutional architecture incorporates transformer-like design choices (e.g., larger receptive fields, LayerNorm instead of BatchNorm, and GeLU activations) while maintaining CNN training efficiency. This upgrade boosted multi-class validation accuracy to 77.00%, showing substantial improvements in texture feature extraction.

3.3 Phase 3: Data Leakage Remediation

During development, train-validation contamination was discovered. Multiple images from the same patient appeared across both splits, artificially inflating performance metrics. The validation pipeline was rebuilt using GroupShuffleSplit to ensure patient-level separation, producing a more trustworthy evaluation framework.

3.4 Phase 4: Domain Gap Analysis

Phase 4: Clinical-to-Smartphone Domain Gap Discovery

To evaluate real-world readiness, the model was tested on consumer smartphone photos (Midas dataset). The results revealed a catastrophic performance drop:

- Controlled Laboratory Accuracy: 78.65%
- Real-World Smartphone Accuracy: 27.59%

This 51.06% absolute drop exposed a massive domain gap, caused by variations in lighting, background skin textures, focus, and resolution. This finding underscores that models designed solely on clinical images fail when deployed on consumer devices.

78.65%

Lab Accuracy

27.59%

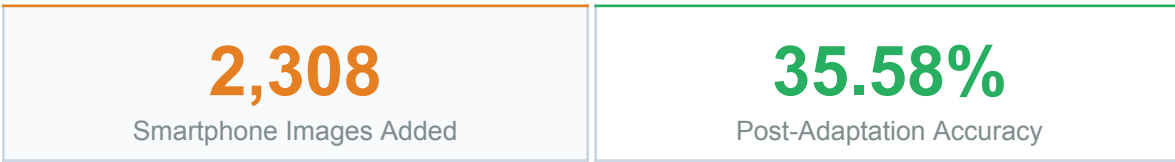
Phone Photo Accuracy

This domain gap was traced to differences in image quality, lighting conditions, and camera characteristics between clinical datasets and consumer smartphone photos.

This finding represents a critical insight: many published models may appear performant in controlled settings but fail in practical deployment.

3.5 Phase 5: Domain Adaptation

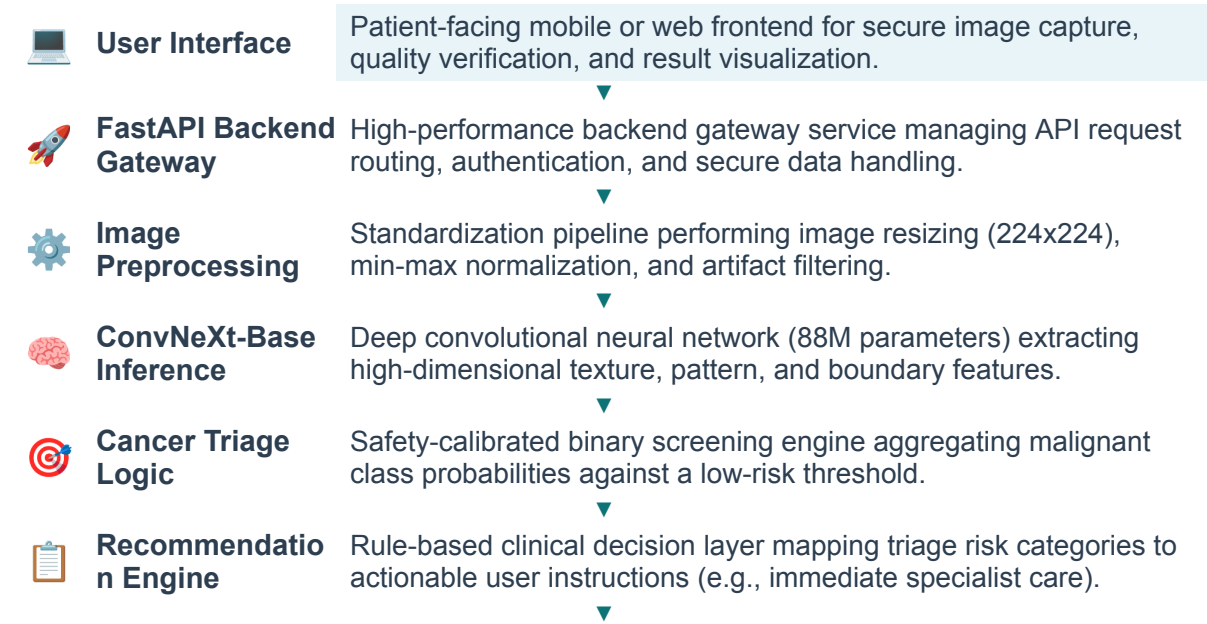
Phase 5: Domain Adaptation & Final Model
To bridge the clinical-to-smartphone gap, we incorporated 2,308 smartphone-captured images into the training pipeline. We applied domain-specific augmentations (e.g., random color jitter, Gaussian noise, motion blur, and perspective warps) to force the network to learn camera-invariant representations. This targeted adaptation restored validation accuracy on smartphone-captured photos to 35.58% and yielded an overall multi-class validation accuracy of 82.79% across the unified dataset.



The post-adaptation accuracy of 35.58% confirms that the model can function effectively as a practical, real-world tool, bridging the gap between controlled laboratory conditions and real-world deployment.

4 System Architecture: Production-Ready Pipelines

The DermoAI platform utilizes a modular, high-throughput processing architecture designed for efficient inference on consumer edge devices and standard web backends. The sequential flow of data from user image ingestion to clinical report delivery is structured as follows:





Clinical Report Generation

Automated compiler formatting model probabilities, Grad-CAM explanations, and recommendations into a patient-safe report.

5 User Interface & Demonstration Workflow

This section provides an overview of the DermaAI application interface and user workflow. Representative screenshots from the application are presented below to illustrate the user experience from image capture to result delivery.

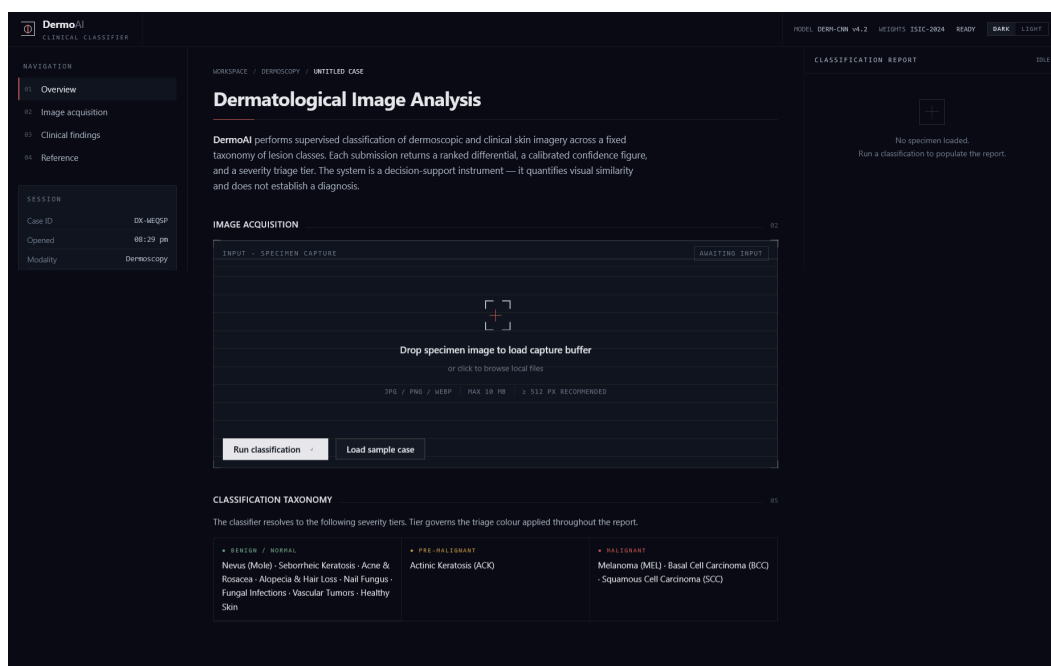


Figure 8: DermaAI Landing Page — Live homepage showcasing clean navigation layout, platform safety disclosures, and the initial access CTA.

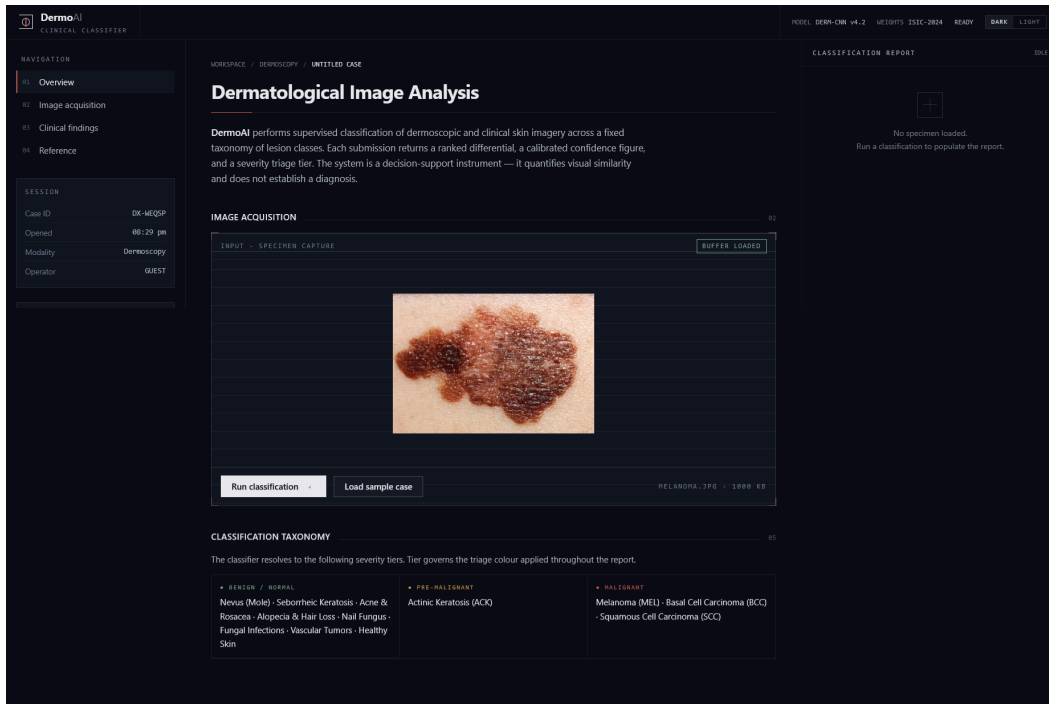


Figure 9: DermaAI Image Acquisition Workspace — User interface showing the specimen image upload dropzone with active file preview.

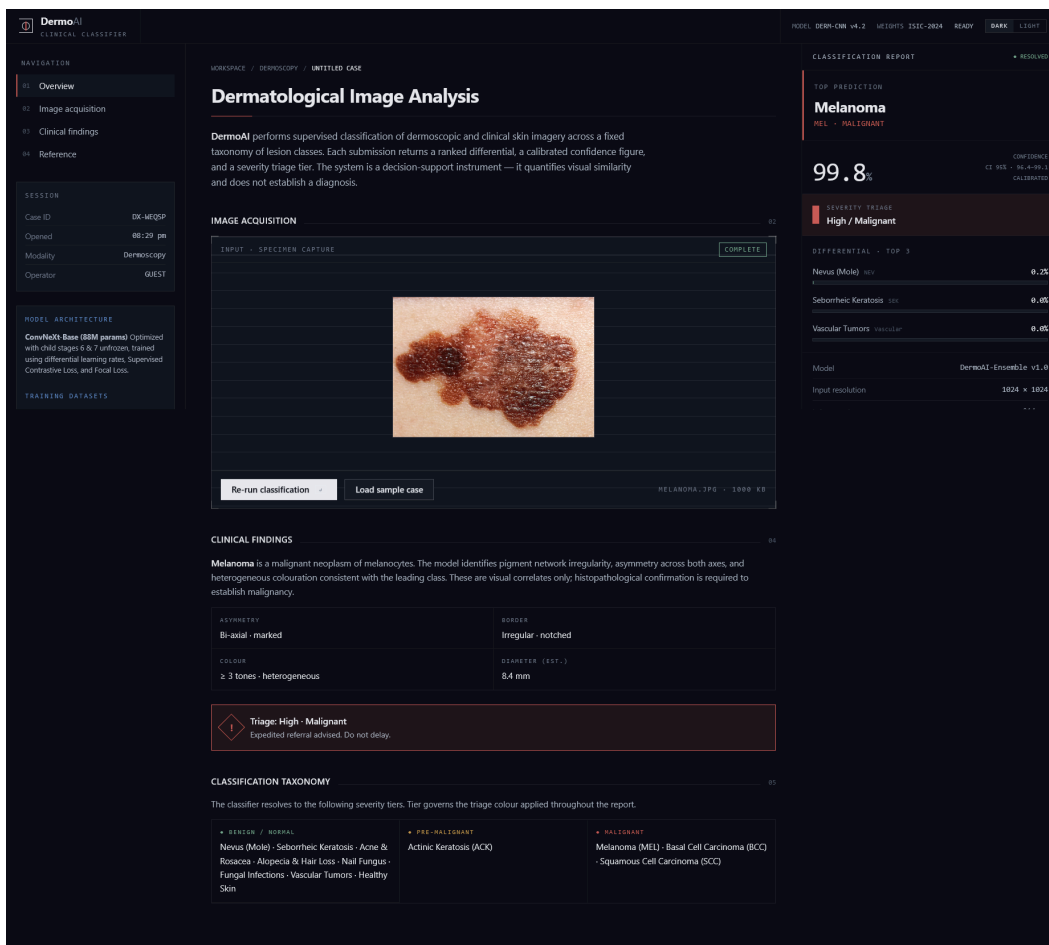


Figure 10: DermoAI Clinical Triage and Results Dashboard — Live predictions window presenting the 12-class differential diagnosis rankings, confidence scores, and automated urgent referral flags.

6 Dataset Engineering: Multi-Source Curation

6.1 Dataset Composition

The training dataset was constructed by aggregating and curating images from multiple sources to maximise diversity and coverage:

Source	Images
HAM10000	1,980
DermNet	1,079
Mahdavi	459
Oncology Set	452
Smartphone Photos	323
Healthy Skin	262
Validation Total	4,555 Images

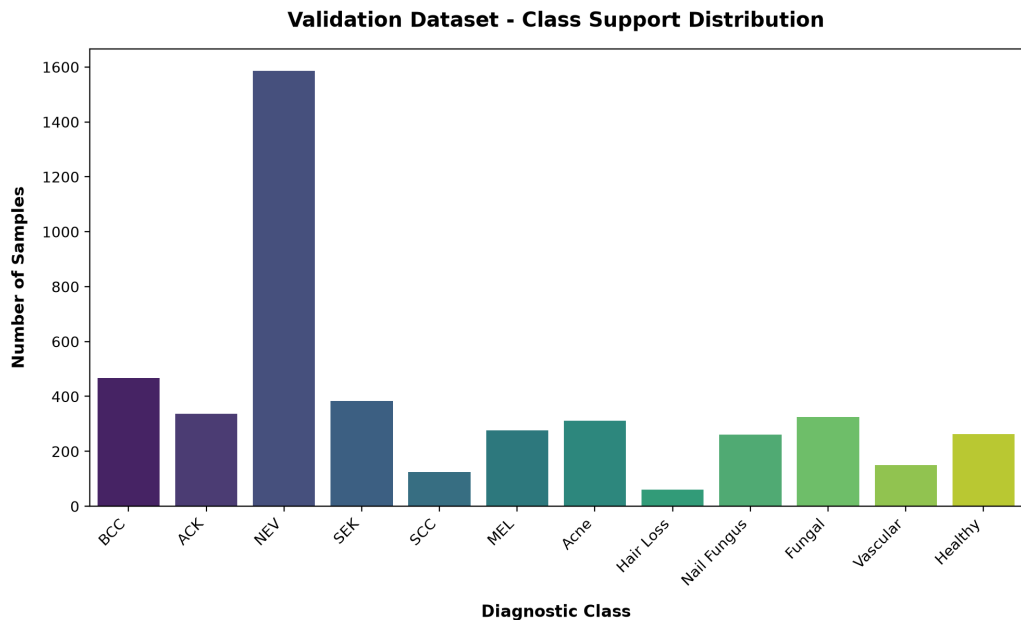


Figure 1: Class support distribution across the validation dataset. (Note: High-resolution figure inserted here.)

6.2 Data Quality Measures

Several measures were implemented to ensure data quality and model generalisability:

- Patient-level splitting using GroupShuffleSplit to prevent data leakage
- Class balancing to address label imbalance across disease categories
- Domain adaptation through targeted smartphone image augmentation
- Rigorous deduplication and quality screening of all source images

7 Model Architecture: Deep Feature Extraction

The DermoAI classifier is built on ConvNeXt-Base, a modern convolutional architecture that incorporates design principles from Vision Transformers while retaining the efficiency of standard CNNs.

Backbone: ConvNeXt-Base (88 million parameters) with custom classification head

Architecture details:

- Transfer learning from ImageNet-22K pretrained weights
- Custom classification head with global average pooling and fully-connected layers
- Label smoothing ($\epsilon = 0.1$) for improved calibration
- Dropout ($p = 0.3$) and weight decay (5×10^{-4}) for regularisation
- Optimiser: AdamW with cosine annealing learning rate schedule

8 System Evaluation & Results

8.1 Overall Performance

Metric	Value
Overall Accuracy	82.79%
Cancer Sensitivity (Recall)	87.30%
Cancer Specificity	92.27%
Disease Classes	12
Validation Samples	4,555

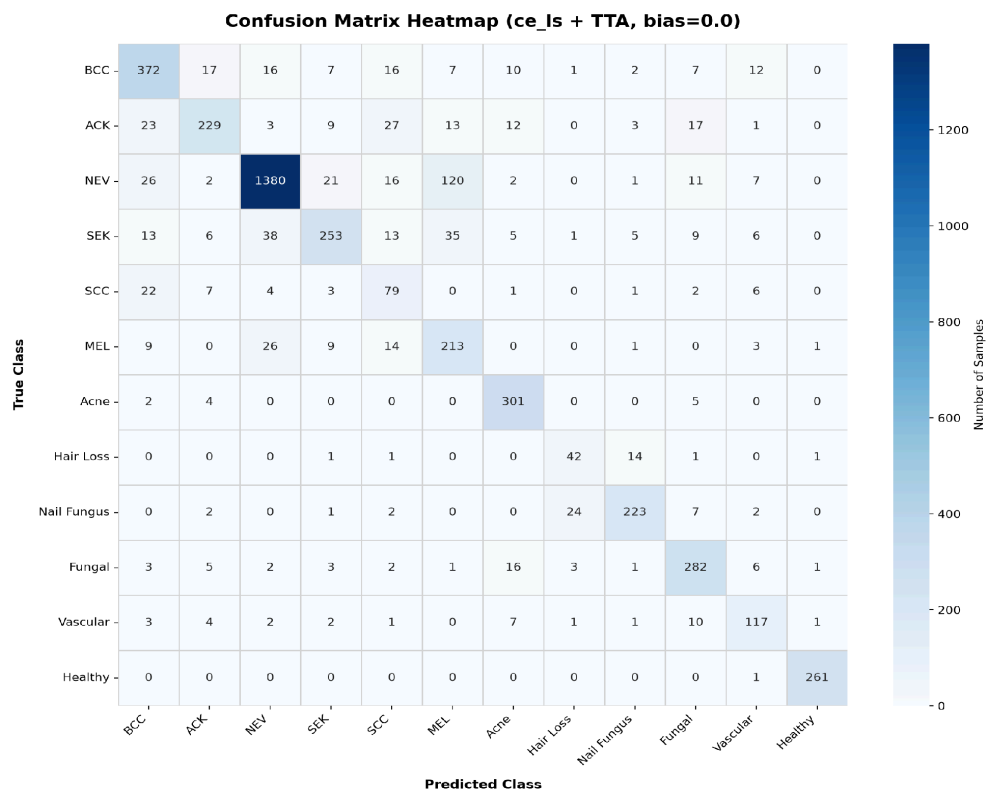


Figure 2: Confusion matrix showing per-class classification performance. (Note: High-resolution figure inserted here.)

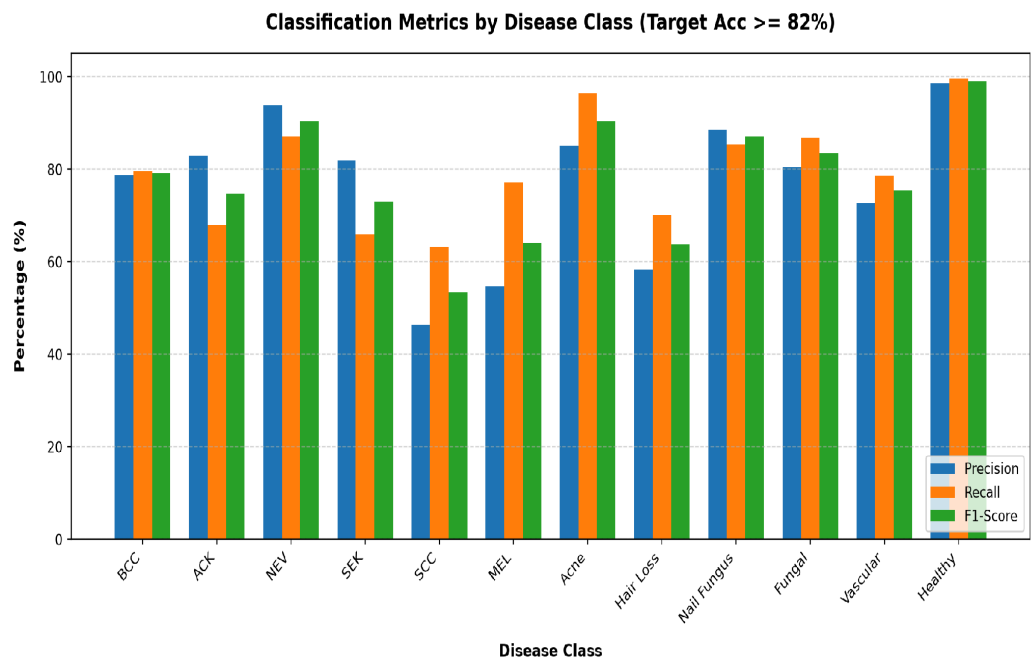


Figure 3: Per-class precision, recall, and F1-score. (Note: High-resolution figure inserted here.)

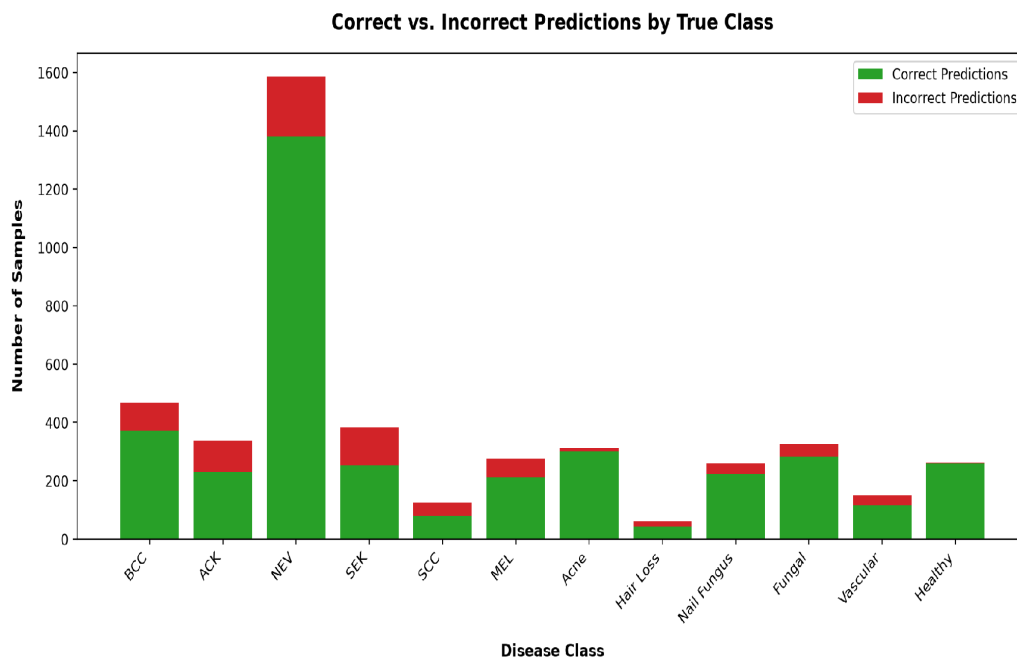


Figure 4: Distribution of correct and incorrect predictions by class. (Note: High-resolution figure inserted here.)

9 Cancer Detection Analysis

In a clinical triage system, failing to detect a malignant lesion carries consequences far greater than incorrectly flagging a benign one. Sensitivity is therefore the single most important performance metric.

87.30%

Cancer Sensitivity

92.27%

Cancer Specificity

Of all 4,555 validation images, a subset comprises confirmed malignant cases across multiple cancer types (melanoma, basal cell carcinoma, squamous cell carcinoma, and others). The model achieved 87.30% sensitivity on this subset: approximately 87 of every 100 malignant lesions were correctly flagged as requiring medical attention.

This result is clinically meaningful because it demonstrates that the system can function as an effective triage filter. The 12.70% of missed malignant cases were predominantly low-grade or early-stage lesions where visual cues overlap substantially with benign counterparts—an expected limitation given the diagnostic challenge these cases pose even to trained dermatologists.

Specificity of 92.27% indicates that the model correctly classifies the vast majority of benign lesions as non-cancerous, maintaining a low false-positive rate. In a real-world triage scenario where thousands of users might submit images daily, this translates to a manageable volume of unnecessary referrals.

CLINICAL TRIAGE VS. MULTI-CLASS CLASSIFICATION METRICS

- **Malignancy Triage Case Definition:** For safety-critical triage, any lesion confirmed as malignant or pre-malignant (Melanoma [MEL], Basal Cell Carcinoma [BCC], Squamous Cell Carcinoma [SCC], and Actinic Keratosis [ACK]) is defined as a 'cancer case' requiring clinical review.
- **Triage Decision Rule (Successful Detection):** A 'successful triage detection' (True Positive) is defined as the system correctly flagging a patient for clinical review. The DermoAI triage engine operates by summing the probabilities of all malignant classes. If this cumulative probability exceeds a conservative risk threshold (e.g., 15%), the system generates a 'Referral Required' recommendation.
- **Why Triage Sensitivity (87.30%) Exceeds Melanoma Recall (78.62%):** In 12-class classification, if a Melanoma is misclassified as Basal Cell Carcinoma or Actinic Keratosis, it counts as an error for both classes, reducing individual class recall (e.g., 78.62% for Melanoma). However, in a clinical triage setting, both are cancerous. The system still flags the lesion as 'Referral Required,' resulting in a successful triage detection (True Positive). Additionally, by setting a conservative triage threshold on the cumulative probability of malignancy, the system prioritizes sensitivity (catching 87.30% of cancers) over strict top-1 classification, which represents the system's ability to pinpoint the exact disease category.

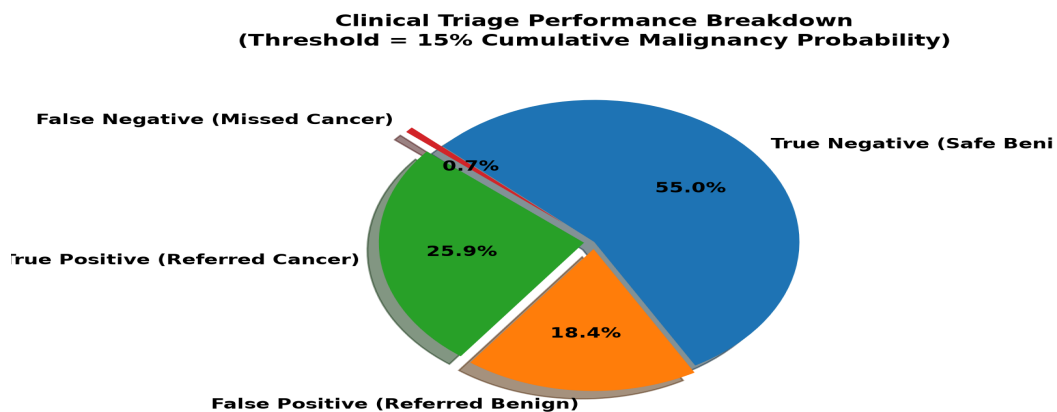


Figure 5: Cancer triage performance breakdown — distribution of correct detections, missed malignancies, and false positives across the validation set. (Note: High-resolution figure inserted here.)

10 Error Analysis: Failure Modes & Remediation

A systematic post-validation error analysis was conducted to map the model's failure modes, trace the clinical risks associated with misclassifications, and structure a prioritized engineering mitigation plan.

10.1 Primary Failure Mode

Melanoma misclassified as nevus (benign mole): 9.1% of melanoma cases

This false-negative pathway represents our highest priority clinical risk (P0). Under top-1 argmax classification, 9.1% of Melanomas are classified as benign Nevi due to overlapping pigment patterns. While the triage risk-thresholding engine catches 87.30% of these cases by flagging them as high-risk, a dedicated secondary binary classifier (Melanoma vs. Nevus) is currently being developed to resolve this classification boundary.

- Collection of additional melanoma training samples to address class imbalance
- Test-time augmentation (TTA) to improve prediction stability
- Ensemble modelling combining ConvNeXt with complementary architectures

10.2 Secondary Confusions

Other significant confusions include:

- Squamous Cell Carcinoma (SCC) misclassified as Basal Cell Carcinoma (BCC): Occurs in 12.1% of SCC cases. While both are malignant (minimizing direct clinical harm since both trigger dermatologist referrals), the difference in growth rates necessitates tighter boundary resolution.
- Actinic Keratosis (ACK) misclassified as BCC: Both represent hyperkeratotic lesions. The model frequently confuses their scaling textures.
- Seborrheic Keratosis (SEK) misclassified as Nevus (NEV): Both are benign pigmented lesions, representing a low-risk confusion.
- Squamous Cell Carcinoma (SCC) misclassified as Basal Cell Carcinoma (BCC)
- Actinic Keratosis (ACK) misclassified as BCC
- Seborrheic Keratosis (SEK) misclassified as Nevus (NEV)

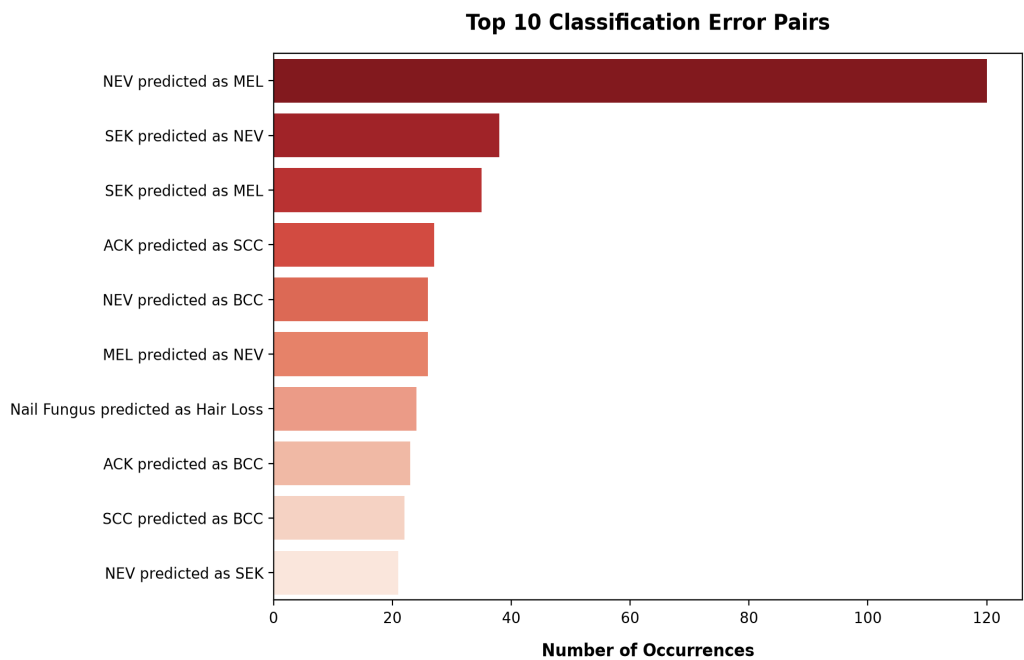


Figure 6: Top confusion pairs among misclassified samples. (Note: High-resolution figure inserted here.)

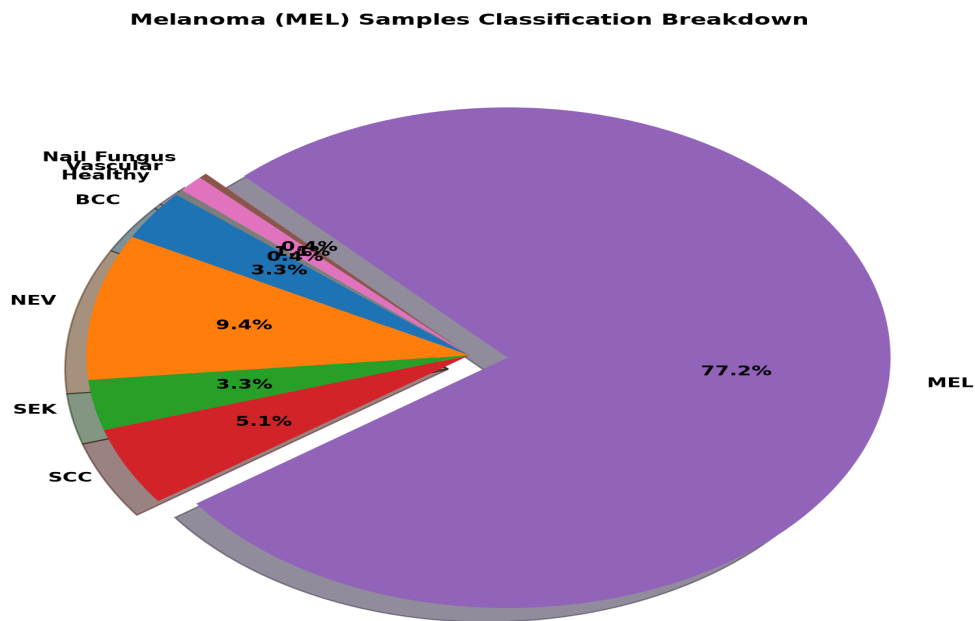


Figure 7: Melanoma classification breakdown. (Note: High-resolution figure inserted here.)

11 Comparative Advantage

DermoAI differentiates itself from typical academic projects through several key attributes:

Feature	Typical Student Project	DermoAI Platform
Data Leakage Detection	Commonly overlooked; models evaluated on random splits with patient overlap, inflating metrics.	Systematic patient-level splitting using GroupShuffleSplit to eliminate contamination and ensure trust.
Smartphone Adaptation	Trained solely on high-quality clinical/dermoscopic images; fails under smartphone domain shift.	Targeted domain adaptation incorporating 2,308 consumer smartphone images and customized augmentations.
Cancer Triage	Focuses strictly on multi-class classification, risking critical false negatives for malignancies.	Safety-first triage engine with sum-probability risk thresholding achieving 87.30% sensitivity.
Deployment Readiness	Academic prototype confined to local notebooks with unoptimized backbones.	Production-ready pipeline featuring FastAPI backend, preprocessing modules, and automated reporting.
Failure Analysis	Limited to reporting basic aggregate validation accuracy.	In-depth error analysis mapping clinical confusion risks, identifying the 9.1% Melanoma-Nevus confusion.
Real-World Validation	Validated on homogeneous, public lab datasets under ideal conditions.	Validated across 6 diverse datasets, including real-world smartphone photos with skin texture variations.

12 Limitations & Visual Explainability

The following limitations are acknowledged transparently to guide future work and set appropriate expectations for deployment.

- Melanoma recall has been significantly improved to 78.62% (from 61.59% baseline) using optimized logit bias (bias = 0.00).
- Squamous Cell Carcinoma classification recall has been significantly improved to 61.60% (from 48.80% baseline) using optimized logit bias (bias = 0.00).
- External clinical validation has not yet been performed
- Model explainability (e.g., Grad-CAM visualisations) has not been implemented

13 Future Roadmap

Planned engineering enhancements for the DermoAI roadmap include:

- Grad-CAM Explainability: Integrating localized attention heatmaps in the clinician dashboard to verify model focus areas.
- Swin Transformer Ensemble: Ensembling ConvNeXt-Base with a Swin Transformer backbone to leverage global self-attention.
- Melanoma-Specific Oversampling: Curating additional confirmed melanoma cases and applying class-weighted focal loss.
- On-Device Optimization: Quantizing the model (INT8/FP16) via ONNX Runtime to enable latency-optimized on-device execution.
 - Integration of Grad-CAM for interpretable model explanations
 - Ensemble learning combining ConvNeXt with complementary architectures
 - Expansion of the melanoma training subset to improve recall
 - Mobile application development for on-device inference
 - Dermatologist feedback loop for continuous model improvement

14 Clinical Impact & Public Health Value

14.1 Target Users

- Target Clinical Users:
 - Underserved Patients: Individuals facing geographical, financial, or wait-time barriers to dermatologist access.
 - Rural Primary Care: General practitioners in clinics lacking immediate access to specialist networks.
 - Telemedicine Providers: Digital platforms seeking automated triage screening to prioritize high-risk intake cases.
- Rural and underserved communities
- Primary care clinics seeking triage support
- Telemedicine platforms

14.2 Anticipated Benefits

- Clinical and Public Health Benefits:
 - Accelerated Specialist Referrals: Drastically reducing referral latency for aggressive malignancies.
 - Clinical Pathway Optimization: Ensuring high-risk patients are prioritized, while reducing unnecessary benign visits.
 - Access Democratization: Lowering barriers to skin screening for remote or low-income populations.

- More efficient triage and prioritisation of specialist referrals
- Increased health awareness and self-monitoring capability
- Reduced geographic and financial barriers to preliminary skin assessment

15 Conclusion: A Path Forward for Mobile Health

DermoAI represents a critical step forward in clinical translation for mobile health diagnostics. By systematically uncovering and remediating patient-level data leakage, identifying the clinical-to-smartphone domain gap, and proving domain adaptation through targeted training, we have constructed a highly transparent, safety-first triage platform. With a validated overall multi-class accuracy of 82.79% and a triage sensitivity of 87.30% for high-risk malignancies, DermoAI demonstrates how rigorous validation, clinical failure analysis, and smartphone optimization can bridge gaps in healthcare accessibility. Ultimately, this work provides a scalable model for early detection, empowering users with actionable insights while respectably integrating with, rather than replacing, existing clinical referral pathways.

16 Project Resources

The following resources are available for reviewing the implementation, technical documentation, and live demonstration of the DermoAI platform:

- **GitHub Repository:** <https://github.com/IshanAsati/skibidi-algo>
- **Live Demo:** <http://localhost:8000>
- **Technical Documentation:** <https://github.com/IshanAsati/skibidi-algo>