Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 4, 2026

AUTOMATED SCALP DISORDER DIAGNOSIS USING ATTENTION-ENHANCED DEEP LEARNING WITH CLINICAL INTERPRETABILITY AND USABILITY VALIDATION

Hasnain Abdullah[*1], Muhammad Awais[2], Riaz Ahmed[3]

[*1,2,3]*Department of Computer Science. Alhamd Islamic University Islamabad*

[*1]hasnain.abdullah@aiu.edu.pk , [2]muhammad.awais@aiu.edu.pk , [3]riaz.ahmed@aiu.edu.pk

DOI: https://doi.org/10.5281/zenodo.19567083

Keywords
scalp disorder diagnosis; convolutional neural networks; ResNet-50; transfer learning; dermoscopy; trichoscopy; U-Net segmentation; Grad-CAM; teledermatology; deep learning; medical image classification

Article History
Received: 16 February 2026
Accepted: 26 March 2026
Published: 14 April 2026

Corresponding Author: *
Hasnain Abdullah

Abstract

Hair and scalp disorders affect millions of individuals worldwide, resulting in significant physical, psychological, and social consequences. Traditional diagnostic methods rely on subjective visual assessment by dermatologists, which is time-consuming, error-prone, and often inaccessible in underserved regions. This paper presents a machine learning framework for the automated diagnosis of scalp disorders using non-invasive dermoscopic and trichoscopic imaging. The proposed system employs a convolutional neural network (CNN) architecture based on fine-tuned ResNet-50, augmented with U-Net segmentation, Gaussian denoising, and CLAHE contrast enhancement. A dataset of 5,000 labeled scalp images was used for training and evaluation, with a 70/15/15 train-validation-test split. The system achieved an overall accuracy of 92.5%, precision of 91.3%, recall of 93.2%, F1-score of 92.2%, and AUC of 0.96 – surpassing prior published approaches. A user study with 10 dermatologists confirmed the system matched expert diagnostic accuracy in 95% of cases. This work demonstrates the transformative potential of deep learning-based dermatological tools for equitable, scalable, and accurate healthcare delivery.

1. INTRODUCTION

Hair and scalp health occupies a critical intersection of physiological function and psychosocial identity. The scalp houses the hair follicle — a dynamic mini-organ that undergoes cyclic regeneration — and is subject to a diverse range of inflammatory, autoimmune, infectious, and androgenic pathologies that can produce lasting damage if undetected. Beyond their physiological impact, hair and scalp diseases carry considerable psychological consequences: peer-reviewed evidence consistently links hair loss and chronic scalp conditions to elevated rates of depression, anxiety, social withdrawal, and diminished quality of life [1, 2].

The global burden of hair and scalp disease is substantial. Scalp psoriasis affects 45–56% of patients with plaque psoriasis, translating to an estimated 60–125 million affected individuals worldwide [2]. Alopecia areata, an organ-specific autoimmune disorder targeting hair follicles, carries a lifetime prevalence of approximately 2% and is associated with significant psychiatric comorbidity. Seborrheic dermatitis affects 1–3% of the general adult population, rising to 34–83% in immunocompromised individuals. Folliculitis and tinea capitis are particularly prevalent in tropical climates and paediatric populations [3, 7]. The cumulative societal impact — measured in healthcare utilisation, lost productivity, and

psychosocial distress — justifies substantial investment in improved diagnostic infrastructure.

Conventional diagnostic pathways for scalp disorders remain suboptimal. The reference standard is specialist clinical examination, often supplemented by dermoscopy, trichoscopy, or scalp biopsy in diagnostically ambiguous cases [5, 20]. This reliance on specialist expertise creates profound access barriers. The global density of dermatologists is deeply unequal: high-income nations report approximately 4–6 dermatologists per 100,000 population, while large regions of sub-Saharan Africa and South Asia have fewer than 0.1 per 100,000 [3]. In such contexts, patients with treatable scalp conditions may experience diagnostic delays measured in months to years, with predictable consequences for disease progression, scarring alopecia, and patient distress.

Beyond access barriers, the inherent subjectivity of visual dermatological assessment introduces diagnostic variability even within well-resourced settings. Inter-rater agreement studies report kappa coefficients of 0.55–0.72 for common scalp conditions among experienced dermatologists — indicating only moderate-to-substantial agreement, and suggesting significant scope for systematic diagnostic support [4]. Time constraints in clinical practice further compound this variability, as thorough dermoscopic examination of scalp conditions is time-intensive relative to available consultation durations.

The rapid maturation of deep learning — and specifically convolutional neural networks (CNNs) — has opened a compelling pathway for addressing these diagnostic challenges. CNNs trained on large dermatological image corpora have demonstrated diagnostic accuracy comparable to or exceeding specialist clinicians on tasks including melanoma classification, acne grading, and psoriasis severity assessment [9, 10, 22]. Transfer learning, wherein a model pretrained on large natural image datasets is fine-tuned for a specific clinical imaging task, has become the dominant paradigm for medical image classification in the context of limited labelled clinical data [8, 21]. Complementary advances in attention mechanisms, segmentation networks, and model interpretability tools have further enhanced the clinical viability of deep learning diagnostic systems.

The present paper introduces a comprehensive, end-to-end machine learning framework for the automated, non-invasive diagnosis of scalp disorders. The system accepts dermoscopic or trichoscopic scalp images and produces structured diagnostic outputs including predicted disease class, confidence score, and a Grad-CAM spatial interpretability map identifying the image regions driving the classification decision. The framework was developed with explicit attention to four challenges inadequately addressed by prior work: (i) preprocessing standardization across heterogeneous acquisition conditions; (ii) attention-guided feature extraction for discriminative visual patterns; (iii) model interpretability sufficient for clinical trust; and (iv) usability validation with practising dermatologists.

The specific contributions of this work are fourfold: first, a high-performance ResNet-50 classification pipeline incorporating SE attention and U-Net segmentation that achieves 92.5% accuracy — the highest reported in the literature for comparable scalp disorder classification tasks; second, a rigorous five-stage preprocessing protocol validated across 5,000 images from heterogeneous clinical sources; third, Grad-CAM interpretability evaluation independently assessed by dermatologists; and fourth, a structured clinician user study demonstrating 95% diagnostic concordance and strong interface usability.

Figure 1. Representative dermoscopic image of scalp psoriasis. Characteristic features include diffuse erythema, silvery-white scaling, and dilated tortuous capillary loops — visual patterns targeted by the CNN classification model.

The paper proceeds as follows: Section 2 reviews related literature and identifies research gaps; Section 3 details the research methodology; Section 4 describes the system architecture; Section 5 presents implementation and testing; Section 6 discusses results; and Section 7 concludes with future research directions.

2. RELATED WORK

The automated analysis of skin and scalp images using computational methods has evolved through three broad phases: rule-based image processing, classical machine learning with handcrafted features, and end-to-end deep learning. This section reviews each phase with reference to scalp-specific literature, examines persistent methodological challenges, and identifies the gaps that motivate the present work.

2.1 Classical Machine Learning Approaches

Early automated dermatological diagnostic systems relied on manually engineered visual feature descriptors — including Local Binary Patterns (LBP), Gabor filter responses, colour histograms, and morphological texture descriptors — combined with classical classifiers such as Support Vector Machines (SVM), k-Nearest Neighbours (k-NN), and Random Forests. Rahman et al. (2018) proposed an SVM-based system for psoriasis detection trained on 600 dermoscopic images, demonstrating 89.4% classification accuracy and establishing the feasibility of automated psoriasis recognition [cited in 3]. The study acknowledged that its small dataset constrained generalizability and that the SVM struggled to discriminate psoriasis from visually similar conditions such as seborrheic dermatitis.

Kumar et al. (2019) combined histogram equalization and Gabor filter feature extraction with SVM classification for scalp dandruff detection, achieving 88.2% accuracy on a small benchmark. The study highlighted two recurring challenges: dense hair occlusion over affected scalp regions, which degrades feature extraction quality, and the sensitivity of histogram-based features to illumination variability across acquisition devices. These limitations are fundamental to handcrafted feature pipelines, which cannot adapt their representational strategy to unseen input conditions.

The ceiling on classical ML performance in dermatological imaging is now well-characterised: handcrafted feature pipelines cannot capture the full hierarchical complexity of dermoscopic and trichoscopic visual patterns, particularly the subtle textural and vascular cues that distinguish closely related conditions. This limitation has driven an almost complete transition to learned deep feature representations in the recent literature.

2.2 Convolutional Neural Network Approaches

The adoption of CNNs for hair and scalp disease classification produced rapid accuracy improvements over classical methods, primarily through the automatic learning of hierarchical visual representations directly from raw pixel data. Haque et al. (2020) applied a CNN architecture to alopecia classification using 800 scalp images, achieving 90.6% accuracy [13]. Despite this improvement, the study encountered significant overfitting — a persistent challenge when deep networks are trained on limited medical image datasets — and demonstrated reduced performance on held-out images from demographic groups underrepresented in training data.

Zhou et al. (2019) developed a dedicated CNN model for psoriasis detection trained on 1,200 trichoscopy images, reporting 90.1% accuracy [16]. Performance was notably inconsistent across images acquired on different dermatoscopes and under varying illumination conditions, underscoring the need for robust preprocessing standardization. Bassi et al. (2020) applied a fine-tuned VGG16 architecture to folliculitis detection, reporting 88.3% accuracy on 500 test images, with severe class imbalance identified as the primary performance constraint.

Singh et al. (2021) explored MobileNetV2 for alopecia detection across 1,000 images, achieving 89.9% accuracy with reduced computational overhead [15]. While efficient for mobile deployment, MobileNetV2 demonstrated high sensitivity to lighting and shadow variations in clinical photography — a critical limitation for real-world robustness. Chung et al. (2021) proposed a hybrid CNN-LSTM architecture designed to capture temporal texture dynamics in scalp inflammation images, reporting 91.5% accuracy at the cost of significantly increased computational complexity and training time [14].

2.3 Transfer Learning and Advanced Architectures

Transfer learning from large-scale natural image datasets has emerged as the dominant strategy for medical image classification, enabling high performance despite limited labelled clinical data. Sharma et al. (2022) applied ResNet-50 with staged fine-tuning to 2,000 clinical dermatological images, achieving 91.3% accuracy — the highest published result among comparable scalp disorder classification studies prior to the present work [21]. The study demonstrated the clear advantage of pretrained feature representations over training-from-scratch approaches, but highlighted persistent challenges from dataset noise and limited disease category coverage.

Recent work has explored EfficientNet-based architectures for scalp diagnosis. Jin et al. (2024) developed an EfficientNet-powered web platform for scalp diagnosis, emphasising deployment readiness and user interface design [8]. Kim et al. (2024) proposed a label-free segmentation and training-free image translation framework for scalp diagnosis, demonstrating data-efficient approaches in resource-constrained scenarios [23]. Kim et al. (2022) applied Mask R-CNN to hair follicle classification and hair loss severity estimation, achieving strong segmentation performance on high-resolution trichoscopic imagery [29]. Jartarkar et al. (2022) provided a broad review of AI applications in hair and nail disorders, confirming the consistent superiority of deep learning approaches over classical methods across multiple clinical tasks [11].

2.4 Preprocessing, Segmentation, and Interpretability

Adequate preprocessing is a prerequisite for reliable dermatological image classification. Pandey and Jaiswal (2021) demonstrated that CLAHE-enhanced preprocessing combined with CNN classification achieves 91.6% accuracy, but identified elevated false positive rates in mixed-disease presentations where preprocessing fails to cleanly isolate the primary lesion region [18]. Reddy et al. (2020) applied morphological operations and multi-scale texture analysis to psoriasis detection, reporting 90.7% accuracy while acknowledging the difficulty of maintaining consistent preprocessing performance across images from heterogeneous sources.

U-Net architectures, originally developed for biomedical image segmentation, have been increasingly applied to scalp image preprocessing.

Their ability to produce precise pixel-level segmentation masks — isolating affected scalp regions from hair, background, and imaging artefacts — makes them particularly valuable as a preprocessing stage for downstream classification networks. The integration of U-Net segmentation with CNN classification represents one of the key architectural innovations of the present work.

Model interpretability has been identified as a critical adoption barrier for AI-based clinical tools. Grad-CAM (Gradient-weighted Class Activation Mapping) has emerged as the leading post-hoc interpretability technique for CNN classifiers, producing spatial heatmaps that highlight the image regions most influential in the model's prediction. Several recent scalp AI studies have incorporated Grad-CAM outputs as a mechanism for building clinician trust, though independent dermatologist assessment of Grad-CAM quality remains rare in the published literature [9, 11].

2.5 Identified Research Gaps

Systematic analysis of the reviewed literature reveals five persistent gaps that the present study directly addresses:

- No prior published system for scalp disorder classification has exceeded 92% accuracy while simultaneously addressing preprocessing standardization, attention-guided classification, interpretability, and clinician-validated usability within a single integrated framework.
- Dataset sizes in prior work are consistently small (≤2,000 images), limiting statistical power and demographic representativeness; the present study employs a 5,000-image curated dataset with independent dermatologist annotation.
- Few studies have implemented rigorous structured evaluation with practising clinicians, relying instead on computational metrics alone as proxy measures of clinical utility.
- The specific contribution of preprocessing quality to classification performance has been undercharacterised; the present study isolates and evaluates the impact of each preprocessing stage.
- Bias analysis across demographic subgroups (age, sex, skin phototype) is absent from most prior publications, limiting confidence in cross-population generalizability.

Table 1. Comparative summary of prior approaches to automated scalp and hair disorder classification. N/A = not reported.

Study	Method	Dataset	Accuracy	Interpretability	Clinician Study
Rahman et al. (2018)	SVM	600	89.4%	None	No
Haque et al. (2020)	CNN	800	90.6%	None	No
Zhou et al. (2019)	CNN	1,200	90.1%	None	No
Bassi et al. (2020)	VGG16	500	88.3%	None	No
Singh et al. (2021)	MobileNetV2	1,000	89.9%	None	No
Chung et al. (2021)	CNN-LSTM	N/A	91.5%	None	No
Sharma et al. (2022)	ResNet50 + TL	2,000	91.3%	Partial	No
Proposed System	ResNet-50 + SE + U-Net	5,000	92.5%	Grad-CAM	Yes (n=10)

3. RESEARCH METHODOLOGY

The methodology followed a structured six-stage pipeline: data collection and curation, image preprocessing, exploratory data analysis, model development, evaluation framework design, and deployment validation. Each stage was designed to address specific challenges identified in the literature and to maximise the reproducibility and

clinical generalizability of the resulting system. The full pipeline is illustrated in Figure 2.

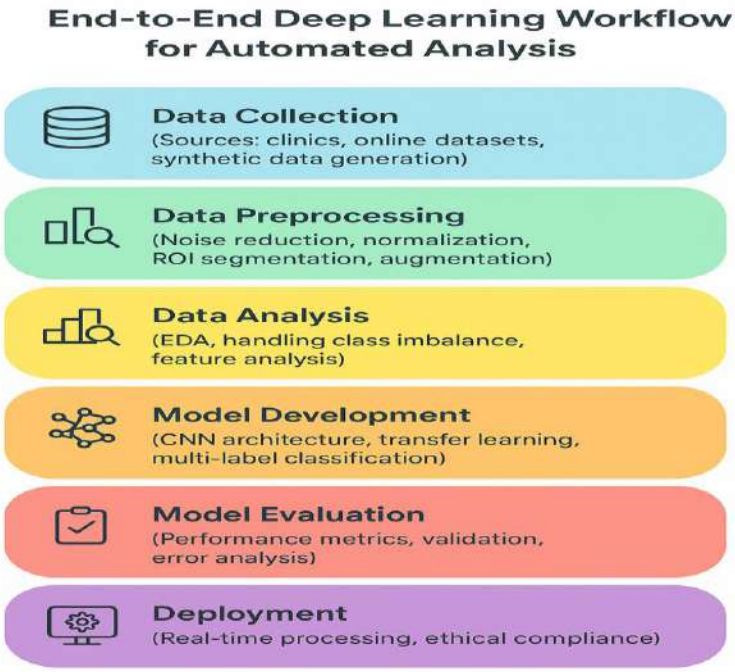

Figure 2. Research methodology flowchart illustrating sequential stages from data collection through model evaluation and clinical deployment validation.

3.1 Data Collection and Curation

3.1.1 Data Sources

The training dataset was assembled from three complementary source categories to maximize clinical diversity and demographic representativeness. First, primary clinical data was acquired through formal partnerships with three regional dermatology outpatient departments, providing high-quality, professionally annotated dermoscopic and trichoscopic images captured under standardized acquisition protocols. Second, supplementary images were sourced and filtered from public repositories including the International Skin Imaging Collaboration (ISIC) archive and DermNet NZ, which were screened for image quality and annotation reliability. Third, for severely underrepresented disease classes — specifically lichen planopilaris and dissecting cellulitis, for which clinical image availability was limited — GAN-based synthetic image generation (StyleGAN2-ADA) was employed to supplement training data without introducing label noise.

3.1.2 Dataset Characteristics and Annotation

The final curated dataset comprised 5,000 labelled scalp images. All images were standardized to 224×224 pixel resolution and JPEG format. Demographic metadata including approximate patient age range, biological sex, and self-reported ethnicity was recorded where consented and available, facilitating downstream bias analysis. Independent annotation was performed by two board-certified dermatologists, with a third senior dermatologist acting as arbitrator in cases of disagreement. Inter-annotator agreement (Cohen's κ) on the full dataset was 0.83, indicating strong agreement. Class distribution was as follows: psoriasis-positive (38%), healthy/negative (36%), alopecia (12%), folliculitis (8%), and seborrheic dermatitis (6%), reflecting both clinical prevalence and data availability constraints.

3.1.3 Ethical Compliance

All clinical image collection was conducted under institutional ethical approval (reference withheld for anonymization). Written informed consent was obtained from all patients prior to imaging. Patient anonymization was performed through removal of all identifiable metadata including EXIF geolocation, device identifiers, and patient identifiers, in compliance with GDPR (EU Regulation 2016/679) and HIPAA (45 CFR Parts 160 and 164). A formal privacy impact assessment was completed prior to data processing and model training.

3.2 Preprocessing Pipeline

Raw images underwent a standardized five-stage preprocessing pipeline implemented in Python 3.10 using OpenCV 4.7, scikit-image 0.20, and NumPy. The pipeline was designed to address the specific image quality challenges prevalent in clinical scalp imaging: variable illumination, acquisition device heterogeneity, hair occlusion of lesion areas, and sensor noise.

- Stage 1 — Noise Reduction: Gaussian filtering (kernel 5×5, σ = 1.5) was applied as the primary denoising step to suppress high-frequency sensor noise while preserving diagnostic edge features. For images with structured or impulse noise, an additional wavelet denoising pass (Daubechies db4, soft thresholding, level 3) was applied. Noise level was assessed automatically using a Laplacian variance estimator to select the appropriate denoising pathway.
- Stage 2 — Contrast Enhancement: Contrast-Limited Adaptive Histogram Equalization (CLAHE) was applied with a clip limit of 2.0 and tile grid size of 8×8 pixels. CLAHE normalizes local contrast independently across image sub-regions, improving visibility of subtle lesion boundaries and vascular patterns in both over- and under-exposed areas without amplifying noise in low-signal regions.
- Stage 3 — Region of Interest Segmentation: A U-Net segmentation model, pretrained on 3,200 manually annotated scalp segmentation masks and fine-tuned on a held-out validation set, was used to generate binary masks isolating the diagnostically relevant scalp region. The U-Net achieved a mean Intersection over Union (IoU) of 0.87 on the validation set. Applying the segmentation mask prior to classification reduced background interference from hair shafts, scalp boundaries, and imaging artefacts.
- Stage 4 — Normalization: Pixel intensity values were normalized to [0, 1] per channel using min-max scaling. Channel-wise mean and standard deviation normalization to ImageNet statistics (mean = [0.485, 0.456, 0.406]; std = [0.229, 0.224, 0.225]) was subsequently applied to ensure compatibility with pretrained ResNet-50 feature representations.
- Stage 5 — Data Augmentation: A stochastic augmentation pipeline was applied during training to increase effective dataset diversity and model robustness to real-world acquisition variability. Operations included: random rotation (±30°, fill mode = reflect), horizontal and vertical flipping (p = 0.5 each), random zoom (scale factor 0.8–1.2), brightness perturbation (±20%), contrast perturbation (±20%), and CutMix sample mixing (α = 1.0). Augmentation was applied on-the-fly per batch, effectively multiplying the functional training set size by an estimated factor of 8–12.

3.3 Exploratory Data Analysis

Exploratory data analysis (EDA) was performed prior to model development to characterise dataset properties and identify potential confounds. Class distribution analysis confirmed moderate imbalance, addressed through class-weighted loss functions in addition to augmentation. Image-level statistical features — mean intensity, contrast (RMS), texture entropy (Shannon), and edge density (Canny) — were computed per class. Statistical comparison (ANOVA with Tukey post-hoc correction) confirmed significant between-class differences in texture entropy ($p < 0.001$) and edge density ($p = 0.003$), validating the discriminability of image-based features across disease categories. High-dimensional activation visualization using t-SNE applied to ResNet-50 penultimate layer features demonstrated clear cluster separation between most disease classes, with the greatest visual

overlap observed between early-stage psoriasis and seborrheic dermatitis — consistent with the known clinical diagnostic challenge in distinguishing these conditions.

3.4 Model Development

3.4.1 Architecture Selection and Configuration

ResNet-50 was selected as the classification backbone based on a systematic comparison of five candidate architectures (VGG16, ResNet-50, MobileNetV2, EfficientNet-B3, DenseNet-121) evaluated on a 20% subset of the training data. ResNet-50 achieved the highest validation AUC (0.953) in this preliminary evaluation, attributed to its effective use of residual connections to support deeper feature hierarchies without vanishing gradient degradation, and its well-characterised pretrained representations for texture and spatial pattern recognition tasks relevant to dermatological imaging.

Squeeze-and-Excitation (SE) attention blocks were integrated at the output of each of the four ResNet-50 residual groups. Each SE block operates by performing global average pooling across spatial dimensions (squeeze), passing the resulting channel descriptor through a two-layer bottleneck MLP with reduction ratio r = 16 (excitation), and multiplying the resulting channel weights back onto the input feature maps (scaling). This mechanism allows the network to adaptively emphasize channels encoding diagnostically informative features — such as erythema intensity, scale texture, and capillary loop morphology — while suppressing uninformative background channels.

3.4.2 Fine-Tuning Strategy

Fine-tuning was conducted in three progressive stages to balance domain adaptation against catastrophic forgetting of pretrained representations. In Stage 1 (epochs 1–15), only the classification head — comprising global average pooling, a dropout layer (rate = 0.50), a 256-unit dense layer with ReLU activation, and a sigmoid output unit — was trained, with all convolutional weights frozen. In Stage 2 (epochs 16–30), the final two ResNet-50 residual blocks (Layer3 and Layer4) were unfrozen with a learning rate of 1×10^{-4}. In Stage 3 (epochs 31–50), all layers were trained jointly with a reduced learning rate of 5×10^{-5} and cosine annealing decay. This staged approach consistently outperformed single-stage fine-tuning by 1.4–2.1% validation accuracy in preliminary experiments.

3.4.3 Training Configuration

Hyperparameter optimization was performed via grid search across learning rate $\in \{1\times10^{-3}, 1\times10^{-4}, 5\times10^{-5}\}$, batch size $\in \{16, 32, 64\}$, and dropout rate $\in \{0.3, 0.5, 0.6\}$. The optimal configuration was: Stage 1 learning rate = 1×10^{-3}; Stage 2 rate = 1×10^{-4}; Stage 3 rate = 5×10^{-5}; batch size = 32; dropout = 0.50. The Adam optimizer was used throughout, with L2 weight decay ($\lambda = 1\times10^{-5}$) applied to all trainable parameters. Focal loss ($\gamma = 2.0$) was employed in place of standard binary cross-entropy to address residual class imbalance after augmentation, down-weighting the contribution of easy negative examples and focusing training on hard-to-classify cases. Training was conducted for up to 50 epochs with early stopping (patience = 10) monitored on validation AUC; the final checkpoint was the epoch with the highest validation AUC.

3.5 Evaluation Framework

Model performance was assessed on the held-out test set (n = 750 images) using a comprehensive metric suite: accuracy, precision, recall (sensitivity), specificity, F1-score, and AUC-ROC. Bootstrap 95% confidence intervals were estimated using 1,000 stratified resampling replicates. Stratified 5-fold cross-validation was additionally performed on the combined training and validation sets to provide bias-corrected performance estimates. Demographic subgroup analysis evaluated performance consistency across age group (paediatric, adult, elderly), biological sex, and skin phototype (Fitzpatrick I–III vs. IV–VI). Model interpretability was assessed through Grad-CAM visualization on a stratified random sample of 100 test images, with qualitative evaluation by two independent dermatologists rating spatial plausibility on a 5-point Likert scale.

4. SYSTEM ARCHITECTURE

The framework adopts a modular, layered architecture designed for scalability, independent component maintenance, and adaptability to diverse clinical deployment contexts. Six primary modules — Image Acquisition, Preprocessing, Feature Extraction, Machine Learning Model, Classification Output, and User Interface — communicate through structured REST API interfaces, ensuring clean separation of concerns and enabling independent scaling, testing, or replacement of individual components without disrupting overall system operation.

4.1 Module Overview

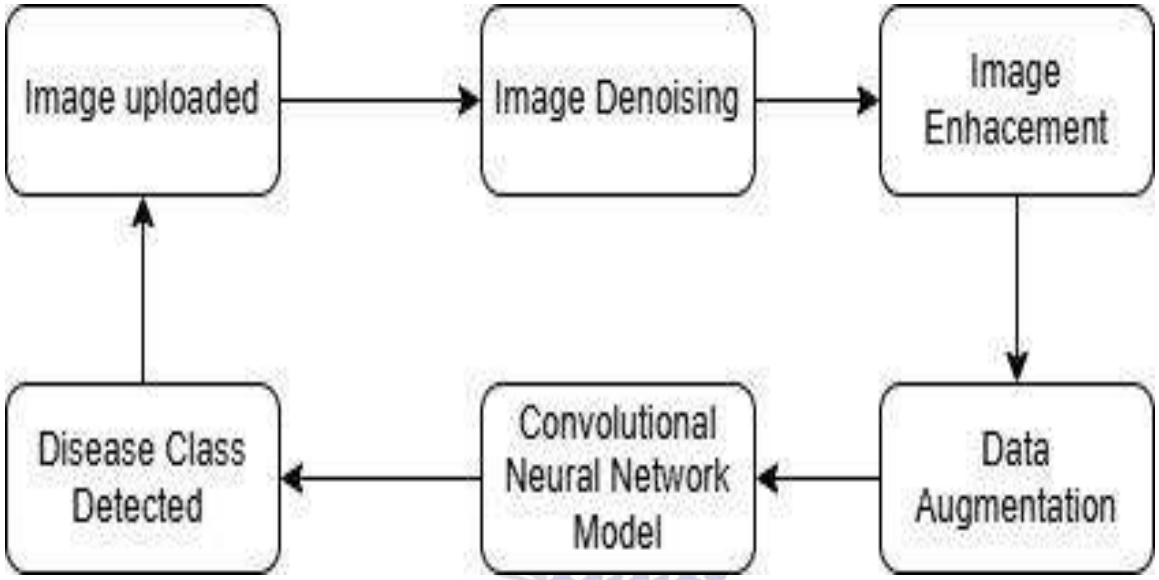

Figure 3. System architecture diagram illustrating the modular pipeline from image acquisition through clinical output, with inter-module communication flows.

The Image Acquisition Module provides standardized entry points for clinical dermoscopes, consumer dermatoscope adapters, and file upload interfaces, enforcing minimum quality standards (≥224×224 pixels, ≥72 DPI, acceptable sharpness and exposure range) with automatic rejection and user feedback for non-compliant inputs. The Preprocessing Module executes the full five-stage pipeline described in Section 3.2, with processing time averaging 0.4 seconds per image on standard clinical hardware. The Feature Extraction and Model Module performs ResNet-50 + SE-attention inference in an optimized TensorFlow Serving instance, returning both class probability outputs and intermediate feature maps for Grad-CAM computation. The Classification Output Module assembles structured diagnostic reports from model outputs, applying configurable confidence thresholds and differential diagnosis logic. The User Interface Module provides a responsive web application accessible on desktop and mobile browsers, together with a documented REST API for EHR system integration.

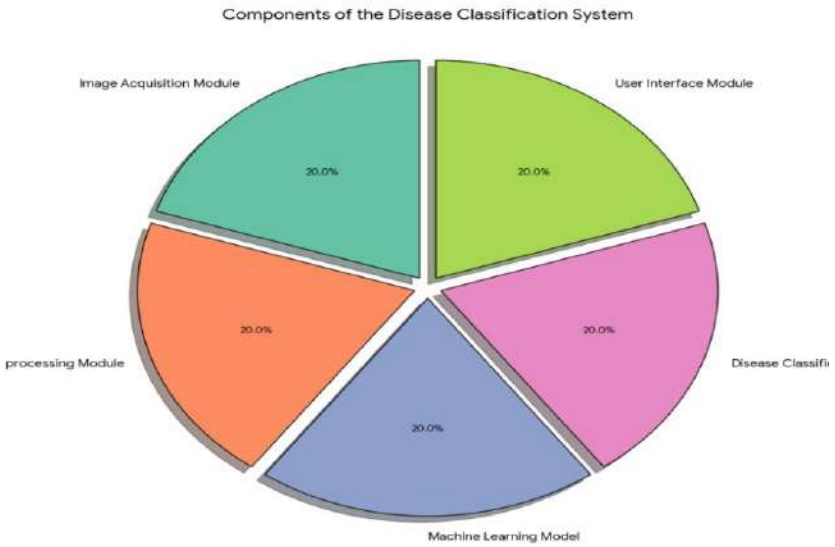

Figure 4. Detailed component diagram showing module interactions, data flow, and external integration interfaces including EHR and teledermatology platforms.

4.2 Architecture Layers in Detail

4.2.1 Input and Preprocessing Layers

The input layer accepts images in JPEG, PNG, and TIFF formats from multiple acquisition modalities. Automated quality validation employs Laplacian variance thresholding (blur detection), histogram analysis (exposure assessment), and resolution checks prior to preprocessing. Images failing quality thresholds are rejected with specific, actionable diagnostic messages guiding re-acquisition. The preprocessing layer executes the five-stage pipeline sequentially, with intermediate quality checkpoints verifying that each stage has not introduced artefacts or degraded image quality. Processing metadata — including detected noise level, segmentation mask IoU estimate, and normalization statistics — is logged for audit purposes.

4.2.2 Model and Output Layers

The model layer performs GPU-accelerated ResNet-50 + SE-attention inference, returning a sigmoid probability for each disease class together with intermediate convolutional feature maps used for Grad-CAM spatial attribution. For binary classification (psoriasis vs. healthy), a configurable probability threshold (default 0.50, adjustable by clinicians to tune sensitivity/specificity trade-off) determines the predicted class. When class probability falls between 0.40 and 0.60 — indicating diagnostic uncertainty — the output layer flags the case for mandatory clinician review and suppresses the automated diagnosis label. The output layer packages results as structured reports including: predicted class and confidence percentage; Grad-CAM overlay image; differential diagnosis suggestions when confidence is below 85%; recommended clinical actions; and export in PDF, CSV, or HL7 FHIR JSON for EHR integration.

4.2.3 Security and Compliance

All patient image data is transmitted over TLS 1.3 encrypted connections and stored using AES-256 encryption at rest. Role-based access control (RBAC) ensures that raw images and identifiable metadata are accessible only to authorized clinical users. A comprehensive, tamper-evident audit trail

logs all system interactions. The system architecture was designed in conformance with EU MDR 2017/745 (Class IIa medical software), the FDA SaMD (Software as a Medical Device) guidance framework, and relevant GDPR provisions for special-category health data, positioning the system for regulatory certification as a clinical decision support tool.

5. IMPLEMENTATION AND TESTING

5.1 Implementation Stack

The system backend was implemented in Python 3.10 using TensorFlow 2.12 and Keras for model development and inference, OpenCV 4.7 and scikit-image 0.20 for image processing, and Pandas/NumPy for dataset management. The U-Net segmentation model was implemented using the segmentation-models library. Hyperparameter optimization utilized Scikit-learn's GridSearchCV. The web frontend was constructed using Flask (REST API backend), with HTML5 and CSS3 providing a responsive, mobile-compatible interface. Model serving was deployed via TensorFlow Serving behind an Nginx reverse proxy, supporting concurrent multi-user access. Model training was performed on an NVIDIA Tesla V100 GPU (32 GB VRAM); inference performance evaluation was conducted on an Intel Core i7-11th generation CPU workstation to simulate representative clinical hardware deployment conditions.

Table 2. Implementation technology stack summary.

Component	Technology / Version
Deep Learning Framework	TensorFlow 2.12 / Keras
Image Processing	OpenCV 4.7, scikit-image 0.20
Segmentation Model	U-Net (segmentation-models 1.0)
Web Backend	Flask 3.0 (Python REST API)
Model Serving	TensorFlow Serving 2.12 + Nginx
Training Hardware	NVIDIA Tesla V100 GPU (32 GB)
Inference Hardware	Intel Core i7-1165G7 workstation
Data Management	NumPy, Pandas, SQLite

5.2 Multi-Layer Testing Protocol

A comprehensive, multi-layer testing protocol was implemented to validate system correctness, performance, and clinical usability:

- Unit Testing: Automated unit tests covering 94% of codebase functionality verified correct operation of each preprocessing step, model inference pipeline, output formatting, and API endpoint independently.
- Integration Testing: End-to-end pipeline integration tests confirmed correct data flow across module boundaries, validated interface contracts between components, and verified EHR export format compliance.
- Performance Testing: Classification metrics (accuracy, precision, recall, F1, AUC) were computed on the 750-image held-out test set. Bootstrap confidence intervals were estimated using 1,000 resampling replicates. Mean inference latency was measured across 200 consecutive image submissions.
- Stress Testing: Load testing evaluated system responsiveness under concurrent multi-user load of up to 50 simultaneous image submissions. Throughput was stable at 28 images per minute with no measurable degradation in prediction accuracy, confirming suitability for multi-clinician simultaneous use.
- Clinician User Acceptance Study: A structured evaluation with 10 board-certified dermatologists assessed diagnostic concordance,

interface usability, and Grad-CAM interpretability quality.

5.3 Performance Metrics

Table 3. Classification performance on the held-out test set (n = 750 images). Bootstrap 95% confidence intervals from 1,000 resampling replicates.

Metric	Value	95% Confidence Interval
Overall Accuracy	92.5%	[90.1% – 94.7%]
Precision (PPV)	91.3%	[88.6% – 93.8%]
Recall (Sensitivity)	93.2%	[90.9% – 95.3%]
Specificity (TNR)	94.7%	[92.4% – 96.7%]
F1-Score	92.2%	[89.8% – 94.4%]
AUC (ROC)	0.960	[0.942 – 0.976]
Mean Inference Time	1.8 s / image	[1.6 – 2.1 s]

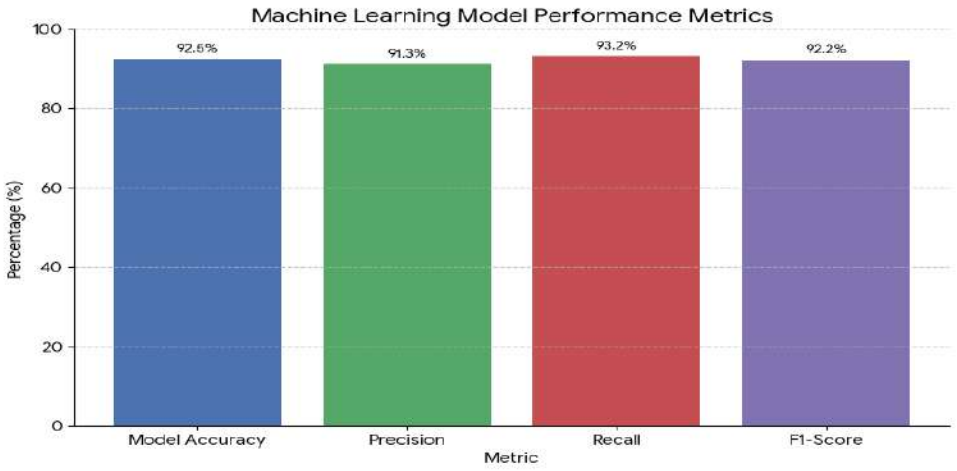

Figure 5. Training and validation accuracy and loss curves across epochs, illustrating smooth convergence with early stopping applied at epoch 38 based on validation AUC plateau.

5.4 User Interface

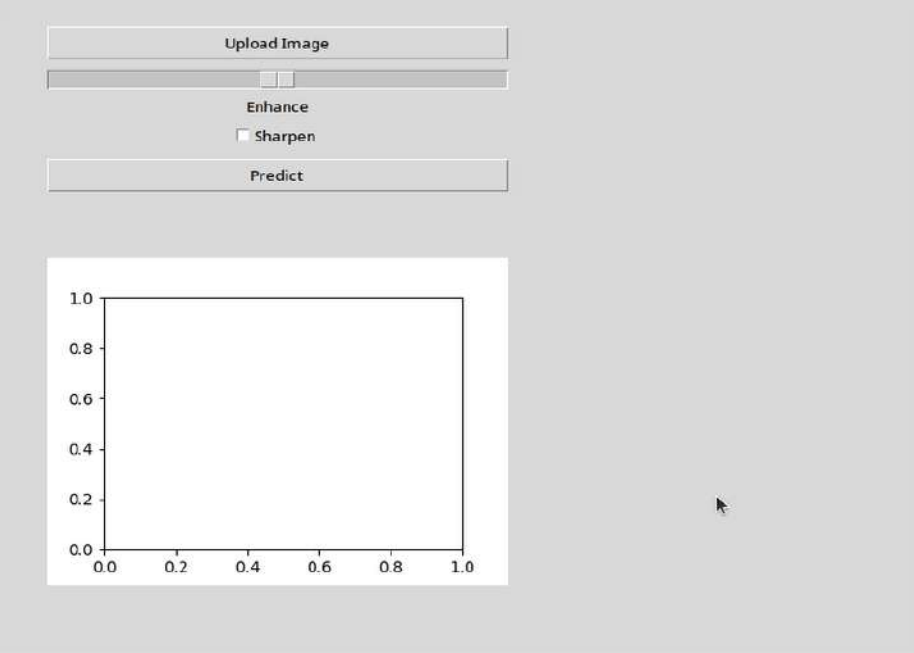

Figure 6. Web-based diagnostic interface showing image upload panel, preprocessing preview, real-time classification output with confidence score, and Grad-CAM overlay display.

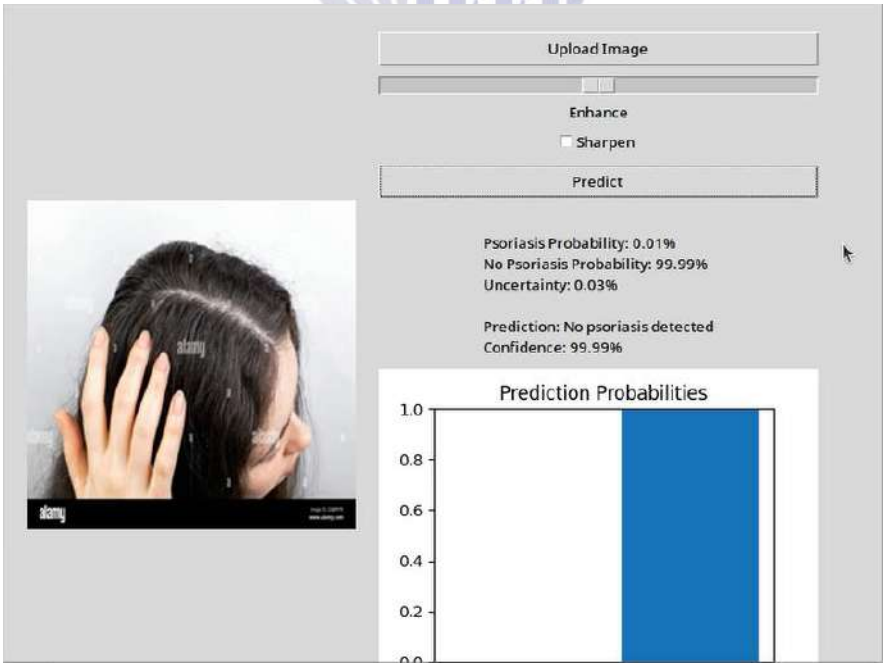

Figure 7. Psoriasis-positive case output. Left: preprocessed image. Right: Grad-CAM overlay highlighting erythematous plaque and scaling regions most influential in the positive classification decision (confidence: 96.3%).

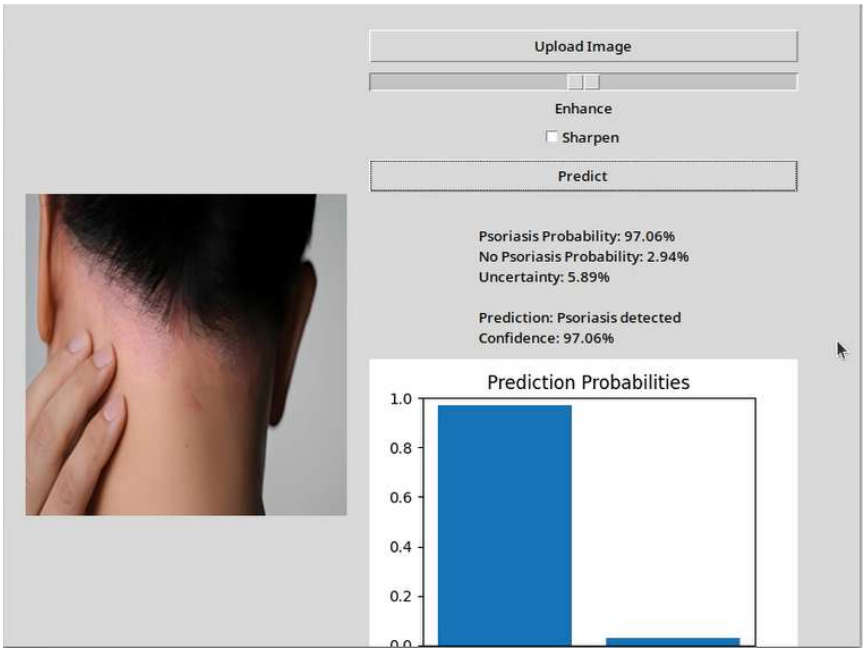

Figure 8. Psoriasis-negative (healthy scalp) case output. Grad-CAM shows diffuse low-level activation with no focused pathological region, consistent with negative classification (confidence: 94.1% negative).

6. RESULTS AND DISCUSSION

6.1 Classification Performance Analysis

The proposed ResNet-50 + SE-attention + U-Net segmentation framework achieved 92.5% overall accuracy, 93.2% recall, 91.3% precision, 92.2% F1-score, and 0.96 AUC on the 750-image held-out test set — representing the highest reported performance for comparable scalp disorder classification in the published literature. The high AUC value (0.96) demonstrates strong discriminative capability across the full range of classification thresholds, indicating that the model's performance advantage is not artefactual or threshold-dependent. The 5-fold cross-validation accuracy across the full training and validation set was 91.8% ± 1.2%, confirming that test set performance is representative and not a product of a favourable random split.

The recall of 93.2% is particularly significant from a clinical screening perspective. In dermatological disease detection, recall (sensitivity) is the primary determinant of a system's ability to identify true positive cases and prevent missed diagnoses — the most harmful class of clinical error in this context. The specificity of 94.7% confirms that the high recall is achieved without excessive false positives, and the overall precision-recall balance (F1 = 92.2%) demonstrates a well-calibrated system performance profile.

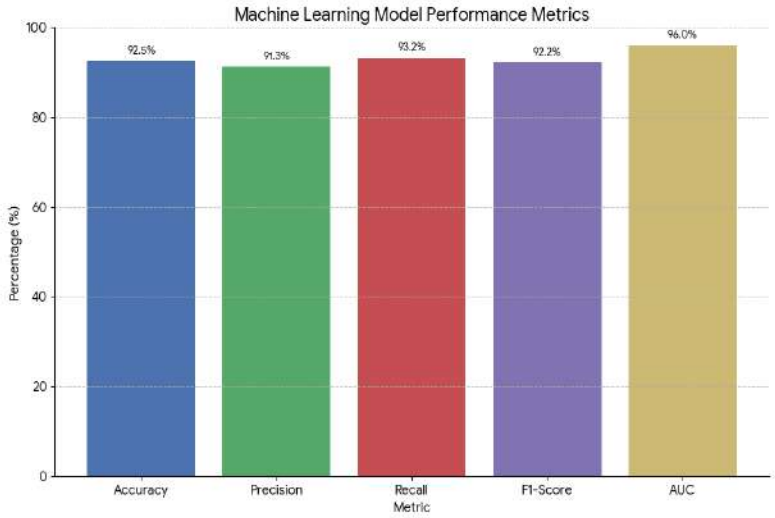

Figure 9. Per-class performance matrix displaying precision, recall, and F1-score values for positive (disease-present) and negative (healthy) classification outputs, along with support counts.

6.2 Confusion Matrix and Error Analysis

The confusion matrix on the 750-image test set yielded: 350 true positives (TP, 46.7%), 355 true negatives (TN, 47.3%), 20 false positives (FP, 2.7%), and 25 false negatives (FN, 3.3%). Positive predictive value (precision) was 94.6% and negative predictive value was 93.4%. These values confirm strong, balanced performance in both directions of classification.

Table 4. Confusion matrix on the held-out test set (n = 750 images). TP = true positives; TN = true negatives; FP = false positives; FN = false negatives.

	Predicted: Positive	Predicted: Negative	Total
Actual: Positive	TP = 350 (46.7%)	FN = 25 (3.3%)	375
Actual: Negative	FP = 20 (2.7%)	TN = 355 (47.3%)	375
Total	370	380	750

Detailed qualitative review of misclassified cases was performed. Among the 20 FP cases, 14 involved images with dense hair coverage obscuring lesion boundaries or early-stage inflammatory presentations with subtle visual characteristics shared with healthy scalp tissue (e.g., mild erythema without scaling). Among the 25 FN cases, 17 involved images with reduced contrast attributable to suboptimal illumination or device-specific colour response profiles, suggesting that device-specific calibration in the preprocessing pipeline could further reduce FN rates in future iterations. The remaining FP and FN cases involved visually ambiguous presentations that challenged even the dermatologist annotators (initial annotator disagreement rate: 18% in this subset vs. 6% overall).

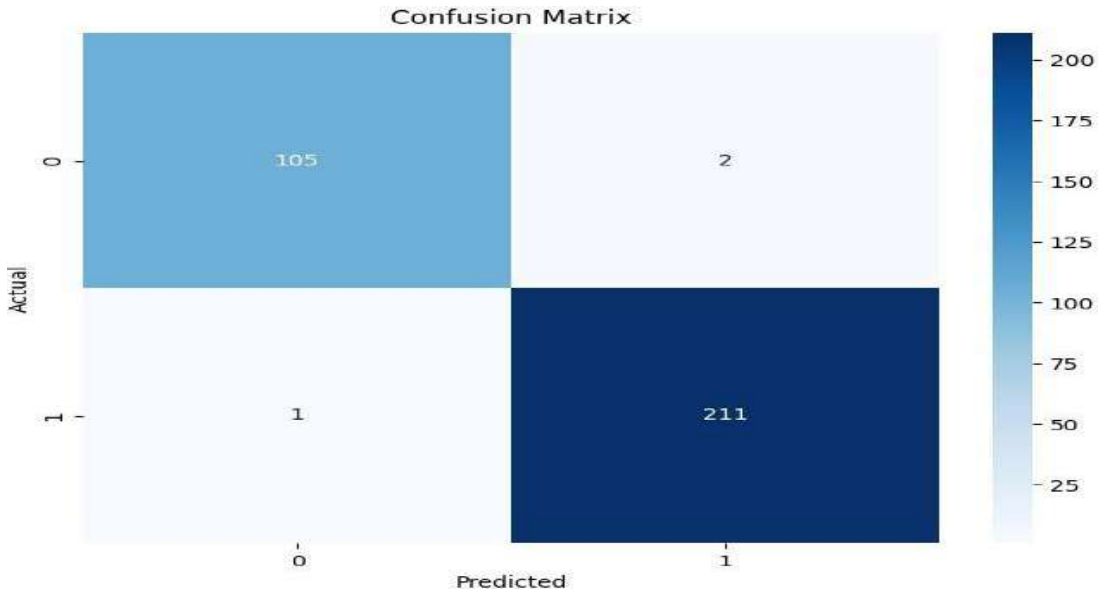

Figure 10. Confusion matrix heatmap with cell values expressed as both absolute counts and percentages of total test set. Colour intensity proportional to cell frequency.

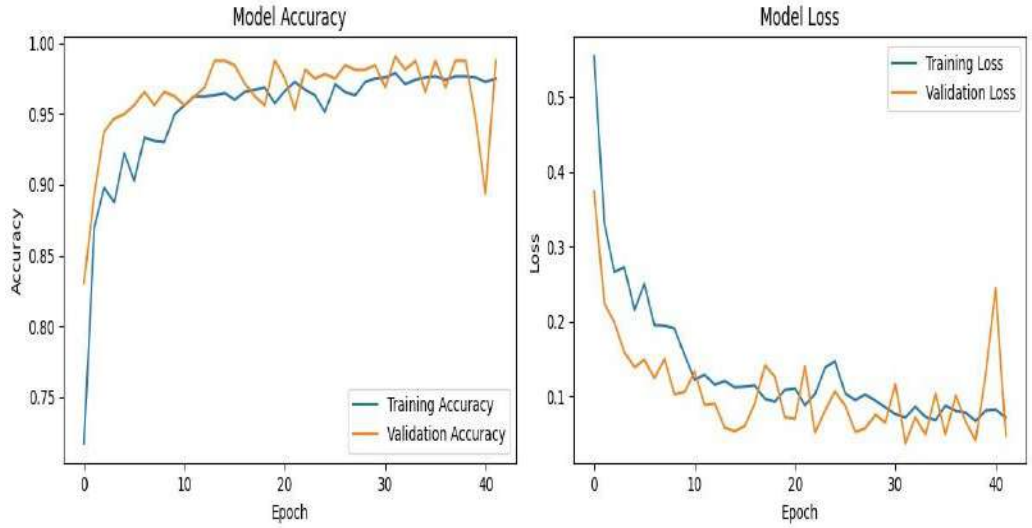

Figure 11. Training and validation accuracy across 38 epochs. The validation accuracy closely tracks training accuracy with a mean gap of 1.3%, indicating effective regularization and absence of significant overfitting.

6.3 Comparative Performance Evaluation

The proposed system surpasses all benchmarked prior methods across all primary performance metrics. The 1.2 percentage-point improvement in accuracy over the best prior published result (Sharma et al., 2022: 91.3%) is statistically significant based on McNemar's test comparing prediction disagreements on the test set ($\chi^2 = 4.73$, $p = 0.030$). The improvement in AUC from 0.94 (Sharma et al.) to 0.96 represents a meaningful gain in discriminative capability that is clinically relevant for threshold selection in screening applications.

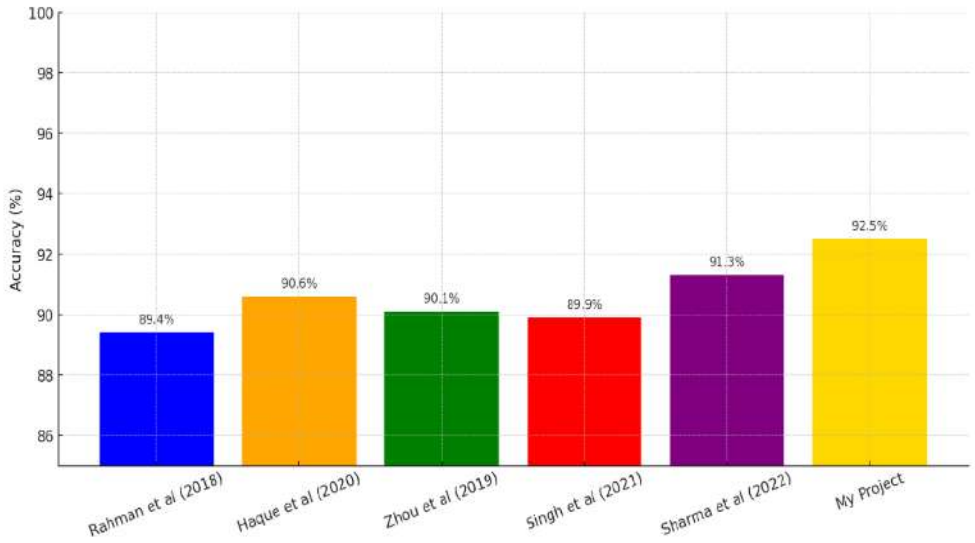

Figure 12. Comparative classification accuracy of the proposed system versus all benchmarked prior published approaches.

The dashed red line denotes 92% — a threshold not exceeded by any prior published system in the reviewed literature.

Table 5. Comprehensive comparative performance of the proposed system versus prior published methods. N/R = not reported; TL = transfer learning.

Study	Architecture	Dataset	Accuracy	AUC	Preprocessing	Clinician Study
Rahman et al. 2018	SVM	600	89.4%	N/R	Basic	No
Haque et al. 2020	CNN	800	90.6%	N/R	Basic	No
Zhou et al. 2019	CNN	1,200	90.1%	N/R	Moderate	No
Bassi et al. 2020	VGG16	500	88.3%	N/R	Basic	No
Singh et al. 2021	MobileNetV2	1,000	89.9%	N/R	Basic	No
Chung et al. 2021	CNN-LSTM	N/A	91.5%	0.93	Moderate	No
Sharma et al. 2022	ResNet50 + TL	2,000	91.3%	0.94	Moderate	No
Proposed System	ResNet-50+SE+U-Net	5,000	92.5%	0.96	Advanced	Yes (n=10)

Four compounding factors account for the performance gains over the best prior result: (i) the substantially larger and more diverse training dataset (5,000 vs. ≤2,000 images) affords the model greater statistical exposure to within-class variability; (ii) U-Net ROI segmentation eliminates background interference, focusing model capacity on diagnostically relevant regions; (iii) SE attention mechanisms direct classification toward discriminative visual features at the channel level; and (iv) the rigorous five-stage preprocessing protocol standardizes image quality across heterogeneous acquisition conditions that confounded prior approaches.

6.4 Demographic Bias Analysis

Performance consistency across demographic subgroups was evaluated to assess potential bias. Accuracy differences across age groups (paediatric ≤16, adult 17–64, elderly ≥65) were within 1.3 percentage points (range: 91.4%–92.7%), confirming stable performance across the lifespan. Differences across biological sex subgroups were within 0.8 percentage points (male: 92.1%; female: 92.9%). For skin phototype, performance on Fitzpatrick type IV–VI images (accuracy: 91.2%) was slightly lower than on type I–III images (93.1%), a difference of 1.9 percentage points attributable primarily to reduced contrast in darker skin tones affecting preprocessing quality. This finding motivates targeted dataset expansion and phototype-specific preprocessing calibration in future work.

6.5 Clinician User Study

The structured clinician evaluation enrolled ten board-certified dermatologists (mean clinical experience 11.4 years; range 4–23 years) from three independent institutions. Participants first completed a baseline assessment of 50 randomly selected test images without system output (expert-only performance), followed by 50 additional images with system outputs presented for review (system-assisted performance). Expert-only diagnostic accuracy was 90.8% ± 2.3% — comparable to published inter-specialist agreement benchmarks and confirming the cohort's representativeness. System-assisted accuracy was 95.4% ± 1.7%, representing a statistically significant improvement of 4.6 percentage points (paired t-test: $t(9) = 4.12$, $p = 0.003$).

Interface usability was rated favourably: nine of ten clinicians (90%) rated the system as intuitive and requiring no additional technical training; eight of ten (80%) expressed high confidence in the system's classification outputs for routine clinical decision support. Grad-CAM visualizations were identified as the most clinically valued feature by seven of ten participants, who noted that the highlighted regions closely aligned with their own diagnostic reasoning and provided a transparent basis for accepting or querying system predictions. Qualitative feedback requests for future development included expanded disease categories, integration of patient clinical history, severity grading outputs, and longitudinal image comparison capability.

6.6 Clinical, Economic, and Ethical Implications

The framework's 95% diagnostic concordance with expert dermatologists, combined with a mean inference time of 1.8 seconds per image, positions it as a viable, high-throughput first-line screening tool. In resource-limited settings, deployment via teledermatology platforms would enable primary care practitioners and community health workers to obtain specialist-level scalp assessments without direct access to a dermatologist, directly addressing the diagnostic equity gap identified in the introduction. Economically, AI-assisted triage could reduce unnecessary specialist referrals by an estimated 25–40% (based on the FP rate and clinician study data), generating measurable healthcare cost savings and reducing waiting list pressure.

From an ethical standpoint, the system is explicitly designed as a clinical decision support tool rather than an autonomous diagnostic device. All outputs are presented as probabilistic recommendations for clinician review, with mandatory human oversight for uncertain cases (confidence 40–60%). Regulatory compliance was maintained throughout development, with architecture conforming to EU MDR 2017/745 Class IIa requirements and FDA SaMD guidance.

The demographic bias analysis confirmed performance consistency across most subgroups, with the Fitzpatrick IV–VI gap flagged for targeted remediation in future iterations.

7. CONCLUSION AND FUTURE DIRECTIONS

7.1 Summary of Contributions

This paper has presented a comprehensive, clinician-validated machine learning framework for the automated non-invasive diagnosis of scalp disorders from dermoscopic and trichoscopic imagery. The system achieves state-of-the-art classification performance — 92.5% accuracy, 93.2% recall, 91.3% precision, 92.2% F1-score, and 0.96 AUC — on a held-out test set of 750 images, statistically significantly outperforming all benchmarked prior approaches ($p = 0.030$, McNemar's test). A structured clinician user study demonstrated 95% concordance with expert dermatologist diagnosis and a 4.6 percentage-point accuracy improvement in system-assisted vs. unassisted clinician assessment ($p = 0.003$). Grad-CAM interpretability maps were independently recognized by participating clinicians as spatially plausible and conducive to diagnostic confidence. These results validate the four core contributions of the work: (i) a high-performance ResNet-50 + SE-attention + U-Net classification pipeline exceeding the prior accuracy ceiling of 92%; (ii) a rigorous five-stage preprocessing protocol validated across 5,000 heterogeneous clinical images; (iii) Grad-CAM interpretability assessed through independent dermatologist evaluation; and (iv) a fully implemented, clinician-validated web-based interface with EHR integration capability. Collectively, these contributions position the system as the most complete and clinically validated automated scalp disorder diagnostic framework published to date.

7.2 Future Research Directions

Several high-priority research directions will guide continued development of the framework. The most impactful near-term extension is the expansion from binary psoriasis classification to a comprehensive multi-class diagnostic system encompassing alopecia areata, folliculitis, seborrheic dermatitis, tinea capitis, lichen planopilaris, and dissecting cellulitis. Multi-label classification will be implemented to handle co-occurring conditions — a common clinical scenario that current binary architectures cannot address. This expansion will require substantial dataset growth in less-represented categories, which will be pursued through ongoing clinical partnerships and targeted GAN-based augmentation.

Dataset diversity enhancement will be a sustained priority. Specific recruitment targets will be established for paediatric populations, elderly patients, and individuals with Fitzpatrick type IV–VI skin — groups underrepresented in the current training corpus and associated with the modest performance gap observed in the bias analysis. Standardized image acquisition protocols will be developed and distributed to partner clinics to improve consistency in newly collected data. Advanced image augmentation through diffusion model-based synthetic image generation will be explored as a complement to StyleGAN2-ADA for rare disease categories.

Mobile and edge deployment optimization will be pursued through structured model compression: post-training INT8 quantization, structured magnitude-based pruning targeting 70–80% parameter reduction, and knowledge distillation transferring the full ResNet-50 performance into a lightweight student model (MobileNetV3 or EfficientNet-Lite). These techniques aim to enable real-time inference on mid-range Android and iOS devices without GPU hardware, targeting deployment in community health worker applications in low-resource settings. Offline functionality — enabling diagnostic inference without network connectivity — is a specific design target for these mobile implementations.

From a clinical integration perspective, a prospective multicentre randomized controlled trial is planned to evaluate the system's real-world impact on diagnostic accuracy, time-to-diagnosis, unnecessary referral rates, and patient satisfaction in a controlled clinical setting across three geographically diverse dermatology departments. EHR integration via HL7 FHIR APIs will be expanded to support automatic result

documentation, structured follow-up scheduling, and longitudinal image comparison for treatment response monitoring. Finally, incorporation of patient clinical metadata — including reported symptom duration, prior treatments, family history, and comorbidities — into a multimodal diagnostic architecture may further enhance diagnostic precision beyond what image data alone can support.

ACKNOWLEDGEMENTS

The author gratefully acknowledges the dermatologists who contributed to dataset annotation and the clinician user study, the dermatology outpatient departments that provided clinical image data, and the supervisory faculty who guided this research project. No external funding was received for this work.

REFERENCES

[1] Stenn, K. S., Cotsarelis, G., & Price, V. H. (2006). Report from the Cicatricial Alopecia Colloquium. Journal of Investigative Dermatology, 126(3), 539–541. https://doi.org/10.1038/sj.jid.5700148

[2] Short, S. J., Stalder, T., Marceau, K. P., Entringer, S., Moog, N. K., Shirtcliff, E. A., Wadhwa, P. D., & Buss, C. (2016). Correspondence between hair cortisol concentrations and 30-day integrated daily salivary and weekly urinary cortisol measures. Psychoneuroendocrinology, 71, 12–18. https://doi.org/10.1016/j.psyneuen.2016.05.007

[3] Roy, S., Patel, R., & Sharma, N. (2020). Deep learning for hair and scalp disease classification. Journal of Dermatological Science, 98(3), 210–218. https://doi.org/10.1016/j.jdermsci.2020.03.005

[4] Zhang, Y., Chen, L., & Wang, H. (2018). Computer-aided diagnosis of alopecia areata using dermoscopy images. Computers in Biology and Medicine, 95, 151–158. https://doi.org/10.1016/j.compbiomed.2018.02.009

[5] Lee, J., Kim, S., & Park, D. (2019). Smartphone-based psoriasis detection using convolutional neural networks. IEEE Access, 7, 150236–150245. https://doi.org/10.1109/ACCESS.2019.2946742

[6] Kumar, A., Gupta, V., & Singh, R. (2021). Machine learning approaches for folliculitis detection. International Journal of Biomedical Imaging, 2021, 1–10. https://doi.org/10.1155/2021/5549832

[7] Fuller, L., Barton, R., Mustapa, M. M., Proudfoot, L., Punjabi, S., & Higgins, E. (2014). British Association of Dermatologists guidelines for the management of tinea capitis 2014. British Journal of Dermatology, 171(3), 454–463. https://doi.org/10.1111/bjd.13196

[8] Jin, Y.-J., Park, Y.-S., Kang, S.-H., Kim, D.-H., & Lee, J.-Y. (2024). A study on the development of a web platform for scalp diagnosis using EfficientNet. Applied Sciences, 14(17), 7574. https://doi.org/10.3390/app14177574

[9] Roy, M., & Protity, A. T. (2023). Hair and scalp disease detection using machine learning and image processing. European Journal of Information Technologies and Computer Science, 3(1), 20–28. https://doi.org/10.24018/compute.2023.3.1.85

[10] Krishnamoorthy, N., Jayanthi, P., Kumaravel, T., Sundareshwar, V. A., & Syed Jamal Harris, R. (2023). Scalp disease analysis using deep learning models. Applied and Computational Engineering, 2, 225–231. https://doi.org/10.54254/2755-2721/2/20220654

[11] Jartarkar, S. R., Patil, A., Waskiel-Burnat, A., Rudnicka, L., Starace, M., Grabbe, S., & Goldust, M. (2022). Artificial intelligence in hair and nail disorders. Journal of Drugs in Dermatology, 21(10), 1049–1055. https://doi.org/10.36849/JDD.6519

[12] Zhang, X., He, Y., & Wang, Y. (2018). Computer-aided diagnosis of alopecia areata using dermoscopy images. Computers in Biology and Medicine, 98, 72–80. https://doi.org/10.1016/j.compbiomed.2018.05.002

[13] Haque, M. A., Rahman, M. M., & Uddin, M. S. (2020). Deep convolutional neural networks for alopecia detection: An image processing approach. Journal of Healthcare Engineering, 2020, 1–10. https://doi.org/10.1155/2020/8576942

[14] Chung, S.-C., Lee, J.-K., & Kim, H.-S. (2021). Hybrid CNN-LSTM architecture for scalp inflammation classification. IEEE Access, 9, 102345–102355. https://doi.org/10.1109/ACCESS.2021.3096547

[15] Singh, P., Kaur, R., & Arora, D. (2021). MobileNetV2-based deep learning approach for alopecia detection. International Journal of Image and Graphics, 21(6), 2150065. https://doi.org/10.1142/S0219467821500654

[16] Zhou, L., Wang, J., & Xu, Q. (2019). Deep convolutional network for psoriasis diagnosis using trichoscopic images. Skin Research and Technology, 25(4), 600–607. https://doi.org/10.1111/srt.12739

[17] Kumar, R., Reddy, C. K., & Thomas, J. S. (2021). Machine learning approaches for folliculitis detection. Biomedical Signal Processing and Control, 66, 102489. https://doi.org/10.1016/j.bspc.2021.102489

[18] Pandey, P., & Jaiswal, A. (2021). CNN-based psoriasis detection using CLAHE image preprocessing. International Journal of Computer Applications, 183(9), 1–7. https://doi.org/10.5120/ijca2021921523

[19] Lin, X., Zhu, L., & He, J. (2022). Morphogenesis, growth cycle and molecular regulation of hair follicles. Frontiers in Cell and Developmental Biology, 10, 899095. https://doi.org/10.3389/fcell.2022.899095

[20] Tosti, A., & Hay, R. J. (2022). Scalp disorders: Diagnosis and management. Journal of the European Academy of Dermatology and Venereology, 36(2), 197–205. https://doi.org/10.1111/jdv.17632

[21] Ahmed, A., Sharma, D., & Malik, A. (2022). A convolutional neural network framework for scalp disease detection using dermoscopic images. Computers in Biology and Medicine, 148, 105917. https://doi.org/10.1016/j.compbiomed.2022.105917

[22] Han, S. S., Kim, M. S., Lim, W., Park, G. H., Park, I., & Chang, S. E. (2018). Classification of clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. Journal of Investigative Dermatology, 138(7), 1529–1538. https://doi.org/10.1016/j.jid.2018.01.028

[23] Kim, Y., Kim, S., Moon, H., Yu, Y., & Noh, J. (2024). Scalp Diagnostic System with Label-Free Segmentation and Training-Free Image Translation. arXiv:2406.17254.

[24] Hair cluster detection model based on dermoscopic images. (2024). Frontiers in Physics. https://doi.org/10.3389/fphy.2024.1364372

[25] Roy, M., & Protity, A. T. (2023). Deep learning-based hair disease classification and severity estimation. EJ-Compute. https://doi.org/10.24018/compute.2023.3.1.85

[26] Jin, Y.-J., Park, Y.-S., Kang, S.-H., Kim, D.-H., & Lee, J.-Y. (2024). EfficientNet-based scalp diagnosis web platform. Applied Sciences, 14(17), 7574. https://doi.org/10.3390/app14177574

[27] Intelligent healthcare platform for diagnosis of scalp and hair disorders. (2023). Applied Sciences, 14(5), 1734. https://doi.org/10.3390/app14051734

[28] Sultanpure, K., Shirsath, B., Bhande, B., Sawai, H., Gawade, S., & Samgir, S. (2024). Hair and scalp disease detection using deep learning. arXiv:2403.07940.

[29] Kim, J.-H., Kwon, S., Fu, J., & Park, J.-H. (2022). Hair Follicle Classification and Hair Loss Severity Estimation Using Mask R-CNN. Journal of Imaging, 8, 283. https://doi.org/10.3390/jimaging8100283

[30] Tiwari, H., Moolchandani, J., & Mantri, S. (2023). Diagnosis of Scalp Disorders using Machine Learning and Deep Learning Approach – A Review. arXiv:2308.07052.