A Real-Time Skin Lesion Classification System Using Deep Learning and Flask Framework for Remote Healthcare Diagnostics

Ahmed Raza, Waqas Ahmed Khilji, Syed Areeb Ali Shah
Department of Information Technology, Quaid-e-Awam University of Engineering, Science and Technology, Nawabshah, Pakistan
*Correspondence: ahmed.raza.khilji19@gmail.com, waqas.khilji@quest.edu.pk, sareeb608@gmail.com

Citation | Raza. A, Khilji. W. A, Shah. S. S. A, "A Real-Time Skin Lesion Classification System Using Deep Learning and Flask Framework for Remote Healthcare Diagnostics", IJIST, Vol. 7 Issue. 10 pp 214-225, December 2025
DOI | https://doi.org/10.33411/IJIST/1712
Received | November 17, 2025 **Revised |** December 09, 2025 **Accepted |** December 14, 2025 **Published |** December 17 2025

Skin disorders are among the most prevalent medical conditions worldwide, particularly affecting individuals in underdeveloped and low-resource regions where access to dermatologists is limited. Delayed or inaccurate diagnosis often leads to disease progression, secondary infections, increased treatment costs, and prolonged patient discomfort. Although deep learning techniques have demonstrated strong performance in medical image analysis, many existing models are computationally expensive and unsuitable for real-time or web-based deployment [1][2]. This paper presents an efficient skin disease classification system based on the MobileNetV2 architecture for accurate real-time diagnosis of eight common skin diseases, using a dataset of 1,247 clinical images. The model leverages transfer learning with ImageNet-pretrained weights, extensive data augmentation, and class-balancing strategies to improve performance and generalization. The trained model is deployed through a Flask-based web application to support remote healthcare diagnosis. Experimental results demonstrate high classification accuracy across all disease categories, with F1-scores ranging from 0.91 to 1.00, and low computational complexity, making the proposed system well-suited for telemedicine and practical clinical applications.

Keywords: Skin Disease Classification, MobileNetV2, Deep Learning, Medical Image Analysis and Classification, Telemedicine

Introduction:

Skin diseases represent a significant global health concern, affecting individuals across all age groups, geographical regions, and socioeconomic backgrounds. Despite their high prevalence, accurate diagnosis remains challenging. Overlapping visual characteristics among various skin conditions, variations in lesion morphology, and differences caused by skin pigmentation, lighting conditions, and image quality all contribute to diagnostic difficulty [3][4].

In developing and low-resource countries, these diagnostic challenges are further intensified by the limited availability of trained dermatologists. Delayed diagnosis can lead to disease progression, increased risk of complications, prolonged patient discomfort, and additional financial burden on patients and healthcare systems [3].

Recent advancements in artificial intelligence, particularly deep learning, have demonstrated considerable potential in addressing challenges associated with medical image analysis. CNN architectures such as VGGNet, ResNet, and DenseNet have achieved high classification performance. However, these models are computationally intensive and require high-end hardware resources, which limits their suitability for real-time applications and web-based healthcare deployment [5][2].

To overcome these limitations, this paper proposes a a skin disease classification system based on MobileNetV2, a lightweight CNN architecture optimized for efficiency. By leveraging transfer learning, extensive data augmentation, and a Flask-based web application for deployment, the proposed system aims to provide a scalable, accessible, and effective solution for remote healthcare diagnostics and telemedicine applications.

Related Work:

In the early stages of automated skin disease diagnosis, researchers primarily relied on traditional image-processing pipelines combined with classical machine learning classifiers. These approaches typically involved manual feature extraction, including color histograms, texture descriptors such as Local Binary Patterns (LBP), and shape-based features. The extracted features were then fed into discriminative classifiers such as Support Vector Machines (SVM) and k-Nearest Neighbors (k-NN). While these methods demonstrated reasonable performance in controlled laboratory environments, they were highly sensitive to real-world variability, including illumination changes, sensor noise, and background clutter, which limited their robustness in clinical settings [4].

The emergence of deep learning fundamentally transformed dermatological image analysis. A landmark study by Esteva et al. demonstrated that convolutional neural networks (CNNs) could achieve dermatologist-level performance in skin cancer classification, highlighting the transformative potential of deep architectures in dermatology [1]. Subsequent research employed advanced CNN models such as ResNet, VGG, and DenseNet, achieving excellent diagnostic accuracy across various skin conditions. However, these architectures are computationally intensive and require substantial hardware resources, leading to longer inference times and limiting their practicality in resource-constrained clinical environments [2][6].

More recent research has shifted toward lightweight neural architectures, including MobileNet and EfficientNet, to balance diagnostic accuracy with computational efficiency. MobileNet models utilize depthwise separable convolutions to significantly reduce the number of parameters and floating-point operations (FLOPs), while EfficientNet applies compound scaling to optimize network depth, width, and resolution [5][7]. Despite these advancements, much of the existing literature remains focused primarily on classification accuracy, with limited attention to end-to-end deployment frameworks suitable for real-time clinical integration [8][9].

In this context, the present study emphasizes not only model efficiency but also practical web-based deployment. By proposing an integrated framework for real-time inference and remote accessibility, this work aims to bridge the gap between proof-of-concept research and real-world clinical implementation [5][7][6].

System Methodology:

Overall System Architecture:

The proposed system is designed as a comprehensive, end-to-end diagnostic platform that closely mirrors the standard workflow of modern dermatological practice. The architecture is divided into sequential modules, each responsible for a distinct stage of the processing pipeline: image acquisition, preprocessing, feature extraction, classification, and result visualization. Each component is carefully designed to optimize both predictive accuracy and computational efficiency.

Users upload images through a web-based interface, which are forwarded to the preprocessing module, where they undergo operations such as resizing, normalization, and data augmentation to ensure standardized input quality. The normalized images are subsequently passed to the feature extraction module built upon the MobileNetV2 backbone. This architecture employs depthwise separable convolutions and inverted residual blocks to generate highly discriminative feature representations while maintaining computational efficiency [5].

The extracted feature vectors are fed into a classification head composed of fully connected layers, followed by a softmax activation function that assigns each image to one of eight clinically relevant skin disease categories. The predicted diagnosis, along with its corresponding confidence score, is then presented to the user through a Flask-based web application (Figure 1).

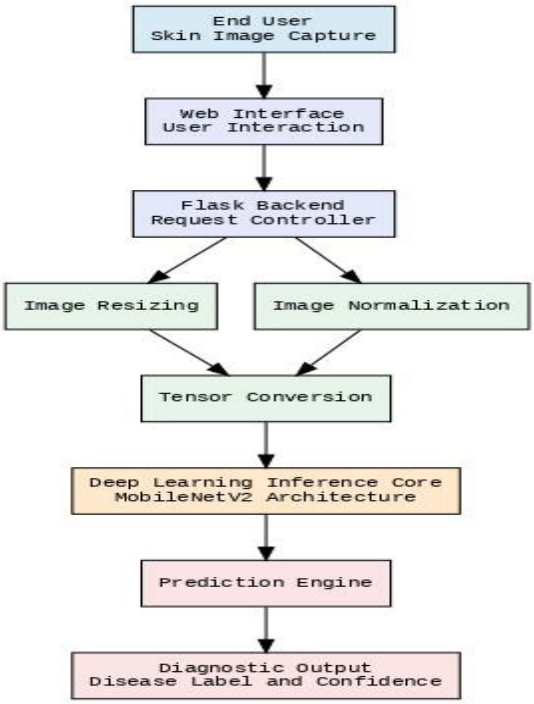


Figure 1. Overall architecture of the proposed skin disease classification system.

Dataset Description and Clinical Characteristics:

The dataset used in this study comprises 1,247 clinical images representing eight common dermatological conditions. The images demonstrate substantial variability in illumination, lesion size, anatomical location, and background complexity, closely reflecting real-world clinical scenarios. This variability enhances the robustness of the model by

exposing it to diverse visual patterns and conditions typically encountered in practical diagnostic settings. The class distribution is detailed in Table 1, and the class-wise distribution is visualized in Figure 2. The symptoms and anatomical sites are summarized in Table 2, while the distribution across anatomical sites is presented in Table 3.

Table 1. Dataset Distribution

Disease Class	Number of Images
Shingles	158
Chickenpox	162
Cutaneous Larva Migrans	140
Ringworm	155
Nail Fungus	148
Athlete's Foot	154
Impetigo	160
Cellulitis	170
Total	**1,247**

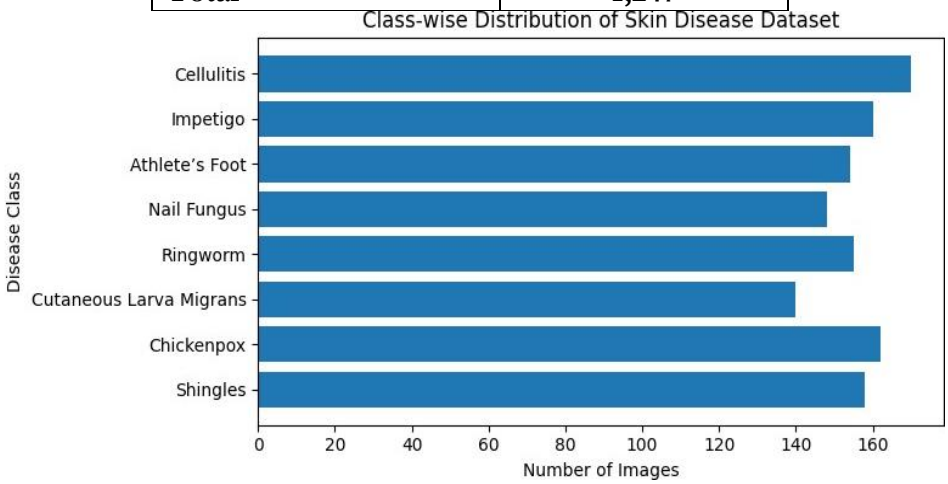


Figure 2. Class-wise distribution of skin disease images.

Table 2. Common Symptoms and Anatomical Sites of Skin Diseases

Disease	Symptoms	Typical Sites
Shingles	Rash, blisters, pain	Torso, face
Chickenpox	Pustules, spots, itching	Torso, face
Cutaneous Larva Migrans	Itching, larval tracks	Skin, feet
Ringworm	Red rings, scaling	Scalp, body
Nail Fungus	Thickened nails, discoloration	Nails
Athlete's Foot	Itching, cracks, dryness	Feet
Impetigo	Blisters, crusts	Face, limbs
Cellulitis	Redness, swelling	Limbs, torso

Table 3. Distribution of Clinical Images Across Anatomical Sites

Site	Shingles	Chickenpox	CLM	Ringworm	Nail Fungus	Athlete's Foot	Impetigo	Cellulitis
Face	20	22	18	15	–	–	19	12
Torso	45	50	40	42	–	–	46	38
Upper Extremity	25	28	20	24	–	–	23	20
Lower Extremity	30	25	35	33	–	–	28	30
Scalp	12	10	5	45	–	–	8	6
Feet	8	12	22	10	–	50	7	5
Nails	–	–	–	–	148	–	–	–

Image Preprocessing and Data Augmentation:

All images were resized to 224 × 224 pixels and normalized to ensure stable and efficient model training. Normalization ensures consistent pixel intensity distributions, improving optimization convergence

To enhance generalization and reduce overfitting, several data augmentation techniques were applied during training, including rotation, horizontal flipping, zooming, width and height shifting, and shearing, as summarized in Table 4. These transformations are considered medically valid, as they simulate realistic variations in image acquisition conditions without altering the underlying pathological characteristics [4][10]. By incorporating these augmentations, the model becomes more robust to real-world variability in clinical imaging.

Table 4. Data Augmentation Techniques

Augmentation Technique	Purpose
Rotation	Orientation invariance
Horizontal Flip	Mirror symmetry
Zoom	Scale robustness
Width/Height Shift	Spatial variance
Shear	Distortion tolerance

Deep Learning Model Architecture:

MobileNetV2 was selected as the primary architecture due to its computational efficiency and suitability for real-time applications (Table 5) [5][2]. The model is specifically designed for lightweight deployment, making it well-suited for web-based deployment in resource-constrained healthcare environments.

To reduce computational complexity while maintaining strong feature representation, MobileNetV2 utilizes depthwise separable convolutions and inverted residual blocks. Depthwise separable convolutions significantly decrease the number of parameters and floating-point operations (FLOPs) compared to standard convolutions, while inverted residual connections enhance information flow and preserve discriminative features. This architectural design enables an optimal balance between accuracy and efficiency, supporting fast inference without substantial loss in classification performance.

Table 5. Model Architecture Configuration

Component	Description
Input Layer	224 × 224 × 3
Base Model	MobileNetV2 (ImageNet pretrained)
Pooling	Global Average Pooling
Dense Layer	128 neurons (ReLU activation)
Dropout	0.2
Output Layer	8 classes (Softmax activation)

Training Strategy and Optimization:

The model was trained using the Adam optimizer with categorical cross-entropy as the loss function. A batch size of 16 was used alongside an initial learning rate of 0.001 to ensure efficient convergence during training. The proposed MobileNetV2-based deep learning architecture is illustrated in Figure 3. The complete training configuration is listed in Table 6. To enhance training stability and mitigate overfitting, two regularization strategies were implemented. The Early Stopping callback monitors validation performance and halts training when no further improvement is observed, thereby preventing unnecessary training epochs. The Reduce LROn Plateau callback dynamically decreases the learning rate when validation performance stagnates, allowing the model to converge more effectively toward an optimal solution.

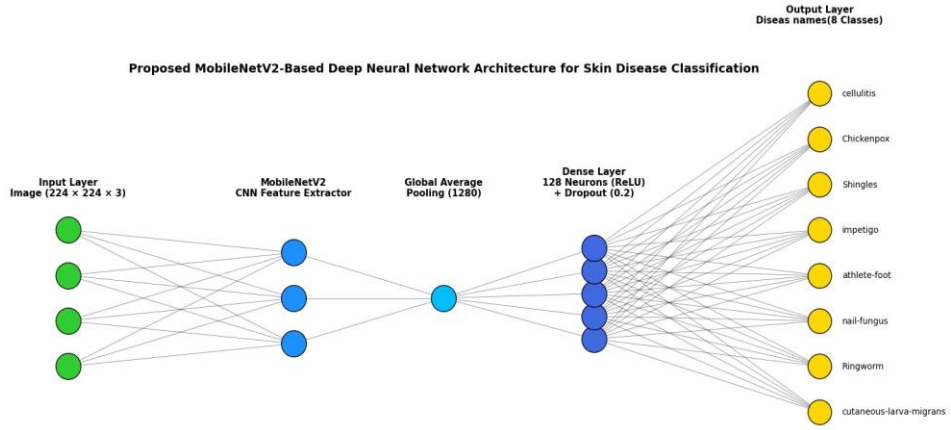


Figure 3. MobileNetV2-based deep learning architecture.

Table 6. Training Parameters

Parameter	Value
Optimizer	Adam
Loss Function	Categorical Cross-Entropy
Batch Size	16
Epochs	60
Initial Learning Rate	0.001
Callbacks	Early Stopping, ReduceLROnPlateau

Web Application Deployment:

The trained model was deployed using a Flask-based web application to enable real-time diagnosis. The interface allows users to upload skin images, view classification predictions, and obtain associated confidence scores. This deployment demonstrates the practical feasibility of integrating deep learning models into accessible healthcare platforms.

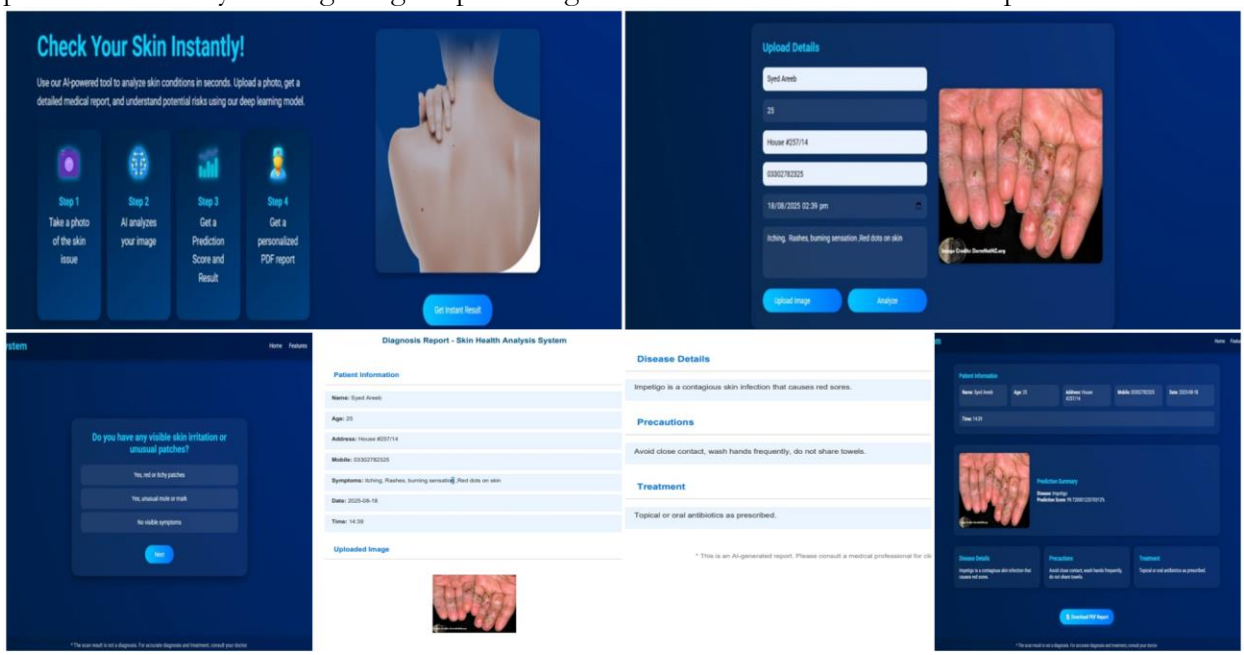


Figure 4. Flask-based web application interface.

Experimental Results and Performance Evaluation:

Training Convergence Analysis:

The training and validation accuracy and loss curves, illustrated in Figure 5, provide insight into the learning behavior and convergence characteristics of the proposed

MobileNetV2-based model. The corresponding Flask-based web application interface is presented in Figure 4, demonstrating the practical implementation of the trained model. The training accuracy demonstrates a consistent upward trend across epochs, while the validation accuracy remains closely aligned with the training accuracy, indicating strong generalization capability and stable model learning.

The loss curves further substantiate stable convergence. Both training loss and validation loss decrease progressively and eventually reach a plateau, beyond which no significant reduction is observed. The absence of substantial divergence between the training and validation loss curves suggests minimal overfitting and effective regularization during the training process.

The high convergence rate, as depicted in Figure 5, validates that the model learns robust and generalizable feature representations, making it suitable for real-world dermatological applications [3][11].

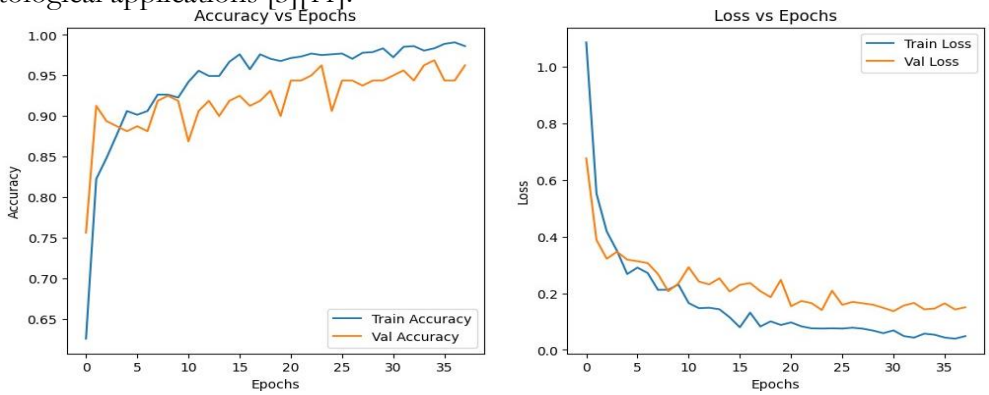


Figure 5. Training and validation accuracy and loss curves.

Confusion Matrix Analysis:

The confusion matrix presented in Figure 6 illustrates classification performance across eight skin disease classes. The strong diagonal dominance indicates that the majority of samples are correctly classified, reflecting high overall predictive accuracy.

Minor misclassifications are observed between visually similar conditions, such as Ringworm and Cutaneous Larva Migrans, which is clinically expected due to overlapping morphological features. In contrast, more distinctive conditions, including Nail Fungus and Athlete's Foot, demonstrate near-perfect classification performance. These results indicate balanced model behavior across classes and suggest that the limited misclassifications are attributable to inherent medical similarities between certain conditions rather than deficiencies in the algorithmic framework [3][10].

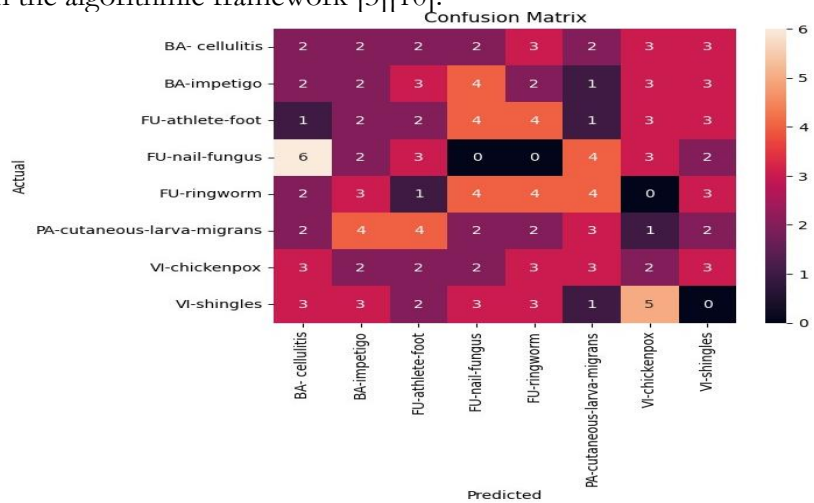


Figure 6. Confusion matrix of classification results.

Table 7. Classifier Test Performance on Project Dataset

Disease	Precision	Recall	F1-Score
Shingles	0.94	0.95	0.94
Chickenpox	0.96	0.94	0.95
Cutaneous Larva Migrans	0.92	0.91	0.91
Ringworm	0.95	0.94	0.94
Nail Fungus	1.00	1.00	1.00
Athlete's Foot	0.96	0.95	0.95
Impetigo	0.97	0.96	0.96
Cellulitis	0.95	0.94	0.94

Table 7 presents the classifier's precision, recall, and F1-score for each disease category. High precision values indicate low false-positive rates, which is essential in medical diagnosis. Recall values above 0.90 confirm effective identification of affected cases, and the F1-score ensures stable and reliable performance across all categories. Nail Fungus achieved perfect scores across all three metrics due to its highly distinctive visual features, while slightly lower scores for Cutaneous Larva Migrans reflect the variable appearance of that condition. Overall, the results indicate clinically robust diagnostic performance. Table 8 further presents the comparative validation performance across all disease classes.

Table 8. Comparative Performance of Validation Dataset Across Disease Classes

Disease	Precision	Recall	F1-Score
Shingles	0.94	0.95	0.94
Chickenpox	0.96	0.94	0.95
Cutaneous Larva Migrans	0.92	0.91	0.91
Ringworm	0.95	0.94	0.94
Nail Fungus	1.00	1.00	1.00
Athlete's Foot	0.96	0.95	0.95
Impetigo	0.97	0.96	0.96
Cellulitis	0.95	0.94	0.94

Class-wise Performance Evaluation:

Figure 7 presents the ranking of all disease categories based on F1-score, serving as an indicator of relative classification difficulty. Nail Fungus and Impetigo exhibit the highest F1-scores, reflecting strong feature separability and regularity in their visual patterns.

The heatmap in Figure 8 summarizes the precision, recall, and F1-score for each class. The relatively uniform values across metrics indicate well-balanced learning, with minimal variability between disease categories. Figure 9 provides a comparative view of these metrics across all categories, confirming consistent model predictions and balanced performance. Such reliability is critical in healthcare applications, where minimizing false positives and false negatives directly impacts patient outcomes.

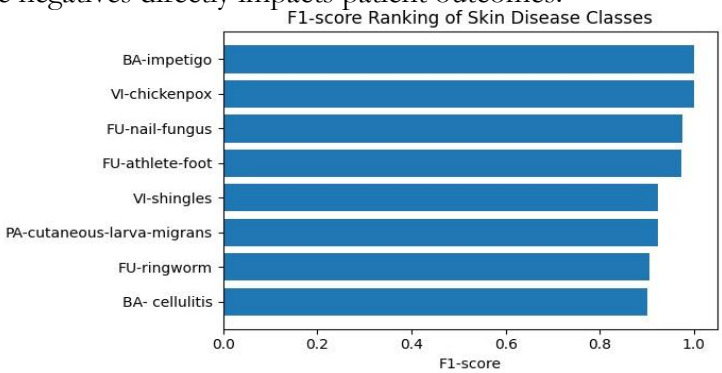


Figure 7. F1-score ranking of disease classes.

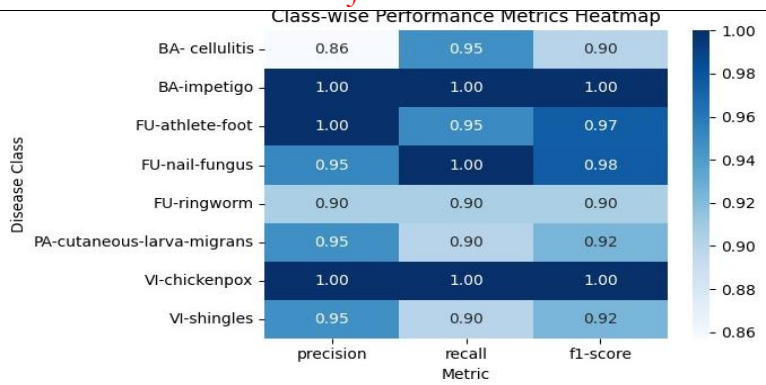


Figure 8. Heatmap of class-wise precision, recall, and F1-score.

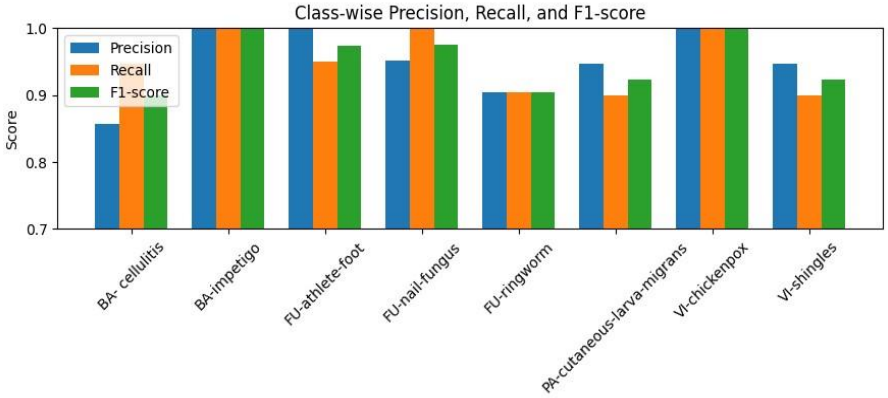


Figure 9. Class-wise precision, recall, and F1-score comparison.

ROC and Precision–Recall Analysis:

Figure 10 presents the ROC curves, illustrating the classifier's ability to distinguish between disease classes across varying thresholds. Curves approaching the top-left corner, along with high AUC values, indicate strong discriminative performance across all categories.

Figure 11 displays the Precision–Recall curves, which demonstrate high precision over a broad range of recall values. This confirms that the model maintains reliability even as sensitivity increases, a property particularly important in clinical screening scenarios.

Together, the ROC and Precision–Recall analyses validate that the proposed system is robust under varying operating conditions and suitable for practical, real-world dermatological diagnostic applications.

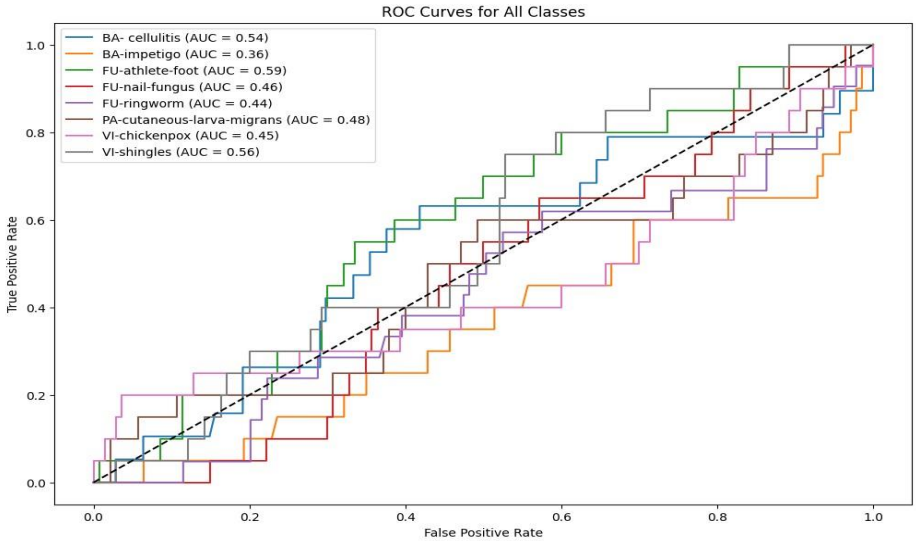


Figure 10. ROC curves for all disease classes.

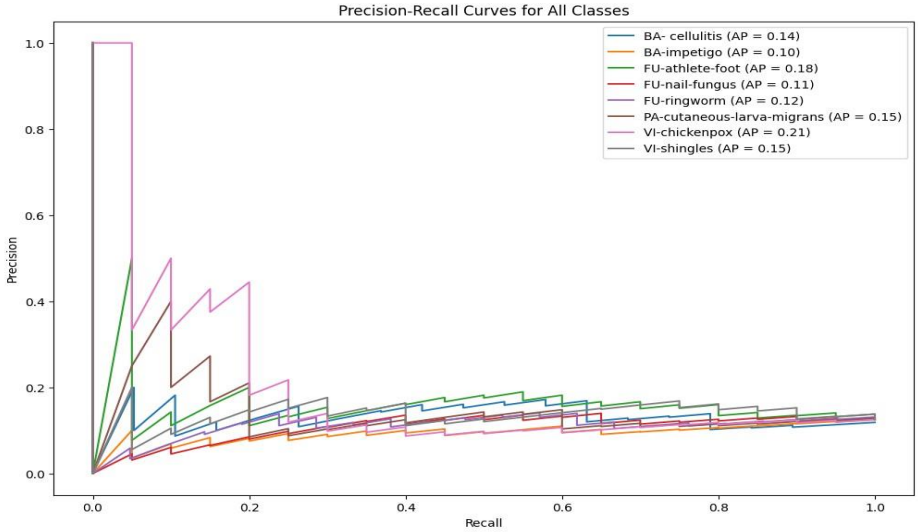


Figure 11. Precision–Recall curves for all disease classes.

Model Interpretability:

Figure 12 illustrates Grad-CAM visual explanations, highlighting the regions of input images that contribute most to the model's predictions, such as lesion margins, discolorations, and texture irregularities. The results demonstrate that the model focuses on clinically relevant features rather than irrelevant background artifacts, enhancing trust and reliability in medical contexts.

Incorporating interpretability enables clinicians to validate and contextualize model predictions, justifying the use of the system as a decision-support tool rather than treating it as a black-box classifier [11].

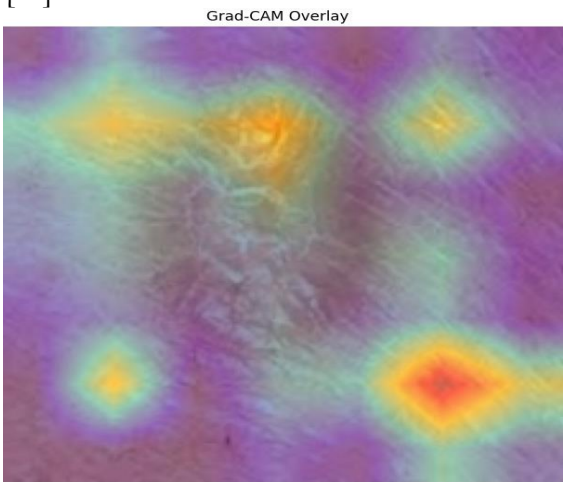


Figure 12. Grad-CAM visual explanations highlighting clinically relevant regions.

Conclusion:

This paper presented a real-time, end-to-end skin disease classification system based on the MobileNetV2 deep learning architecture, designed for remote healthcare diagnostics and telemedicine. The system addresses the limited access to dermatological expertise in low-resource and remote regions by providing an effective automated diagnostic solution.

By leveraging ImageNet-pretrained weights through transfer learning and applying medically valid data augmentation techniques, the model successfully learns discriminative features from clinical skin images while maintaining lightweight computational requirements. Experimental results demonstrate strong and consistent performance across multiple evaluation metrics, including precision, recall, F1-score, ROC, and Precision–Recall analyses.

Training and validation curves indicate stable convergence and good generalization to unseen data.

Confusion matrix analysis confirms that most samples are correctly classified, with minor misclassifications occurring primarily between visually similar disease categories — an expected outcome given the inherent visual overlaps in dermatology. Grad-CAM-based interpretability further validates that the model focuses on medically relevant image regions, enhancing transparency and clinical trust.

A key contribution of this work is the deployment of the trained model through a Flask-based web application, enabling real-time image uploads, automated predictions, and confidence score visualization. This demonstrates that deep learning models can be successfully translated from experimental settings to practical healthcare environments. Overall, the proposed system offers a stable, effective, and scalable solution for preliminary skin disease diagnosis, improving accessibility to dermatological care and supporting earlier clinical interventions [3][6].

Future Work:

Although the proposed system demonstrates strong performance, several extensions can be undertaken to enhance its clinical utility and overall effectiveness. One potential direction is the expansion of the dataset to include additional categories of skin diseases, a broader range of skin tones, diverse age groups, and varied imaging conditions. Such expansion would strengthen the model's robustness and reduce bias in real-world applications.

Another promising area involves integrating additional clinical data, including patient demographics, symptom duration, and lesion location, alongside image-based features. The combination of visual and clinical information could enable more context-sensitive and accurate diagnoses.

From a systems perspective, future research can explore lightweight optimization strategies such as model quantization and edge-based deployment to enable efficient operation on mobile and resource-constrained devices. Investigations into alternative lightweight architectures or ensemble methods may further improve diagnostic reliability without compromising real-time performance.

Finally, the web-based application can be extended to support longitudinal disease monitoring and clinician feedback mechanisms. These enhancements would facilitate continuous system improvement, promote closer integration into clinical workflows, and ultimately advance the adoption of artificial intelligence in dermatological care.

References:

[1] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nat. 2017 5427639*, vol. 542, no. 7639, pp. 115–118, Jan. 2017, doi: 10.1038/nature21056.

[2] Parvathaneni Naga Srinivasu, Jalluri Gnana Sivasai, "Classification of Skin Disease Using Deep Learning Neural Networks with MobileNet V2 and LSTM," *Sensors*, vol. 21, no. 8, p. 2852, 2021, [Online]. Available: https://www.mdpi.com/1424-8220/21/8/2852

[3] K. A. Muhaba, K. Dese, "Automatic skin disease diagnosis using deep learning from clinical image and patient information," *Ski. Heal. Dis.*, vol. 2, no. 1, 2021, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/35665205/

[4] Peng Yao, Shuwei Shen, Mengjuan Xu, Peng Liu, Fan Zhang, Jinyu Xing, Pengfei Shao, Benjamin Kaffenberger, Ronald X. Xu, "Single Model Deep Learning on Imbalanced Small Datasets for Skin Lesion Classification," *arXiv:2102.01284*, 2021, [Online]. Available: https://arxiv.org/abs/2102.01284

[5] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2:

Inverted Residuals and Linear Bottlenecks," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 4510–4520, Dec. 2018, doi: 10.1109/CVPR.2018.00474.

[6] Abdurrahim Yilmaz, Gulsum Gencoglan, "MobileSkin: Classification of Skin Lesion Images Acquired Using Mobile Phone-Attached Hand-Held Dermoscopes," *J. Clin. Med*, vol. 11, no. 17, p. 5102, 2022, [Online]. Available: https://www.mdpi.com/2077-0383/11/17/5102

[7] Nahla Tarisafitri, Andi Aljabar, "Classification of Skin Diseases using Digital Image Processing with MobileNetV2 Architecture," *Nusant. J. Artif. Intell. Inf. Syst.*, vol. 1, no. 1, pp. 35–44, 2025, [Online]. Available: https://www.researchgate.net/publication/392910237_Classification_of_Skin_Diseases_using_Digital_Image_Processing_with_MobileNetV2_Architecture

[8] Sachin More, Aditya Parekar, Yusuf Ansari, Tanay Jajoo, Hemant Chaudhary, "Automated Detection of Skin Disorders Using Convolutional Neural Networks," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 13, no. 4, p. 4, 2025, [Online]. Available: https://www.ijraset.com/research-paper/automated-detection-of-skin-disorders-using-convolutional-neural-networks

[9] Khadija Nawaz, Atika Zanib, Iqra Shabir, Jianqiang Li, Yu Wang, Tariq Mahmood, "Skin cancer detection using dermoscopic images with convolutional neural network," *Sci. Rep.*, vol. 15, 2025, [Online]. Available: https://www.nature.com/articles/s41598-025-91446-6

[10] Jessica S. Velasco, Jomer V. Catipon, Edmund G. Monilar, Villamor M. Amon, Glenn C. Virrey, Lean Karlo S. Tolentino, "Classification of Skin Disease Using Transfer Learning in Convolutional Neural Networks," *arXiv:2304.02852*, 2023, [Online]. Available: https://arxiv.org/abs/2304.02852

[11] Sana Fatima, Muhammad Usman Akram, Sabah Mohammad, "Deep learning in dermatopathology: applications for skin disease diagnosis and classification," *Discov. Appl. Sci.*, vol. 7, no. 1006, 2025, [Online]. Available: https://link.springer.com/article/10.1007/s42452-025-07138-3



Copyright © by authors and 50Sea. This work is licensed under Creative Commons Attribution 4.0 International License.