

A Multi-Model Deep Learning Framework for Automated Dermatological Disease Detection

Rohit Kumar Yadav

Department of Computer Science
and Engineering
Raj Kumar Goel Institute of
Technology Ghaziabad, India

Siddhanth Sharma

Department of Computer Science
and Engineering
Raj Kumar Goel Institute of
Technology Ghaziabad, India

Avnav Teotia

Department of Computer Science
and Engineering
Raj Kumar Goel Institute of
Technology Ghaziabad, India

Prince Kumar Singh

Department of Computer Science
and Engineering
Raj Kumar Goel Institute of
Technology Ghaziabad, India

Ms. Leena Chopra

Assistant Professor
Raj Kumar Goel Institute of
Technology Ghaziabad, India

Abstract—The trend of tele-dermatology services is shifting towards the utilization of images obtained from smartphones and web-based devices. However, there is a significant problem of image quality reduction, blurring, or corruptions, along with the presence of irrelevant content, which may affect the reliability of automated dermatological disease detection. In this context, the current study proposes a multi-stage deep learning model that may improve the reliability of automated dermatological disease classification. The proposed model includes an image integrity validation and blur detection module for image quality enhancement. The binary skin and non-skin image classification module is used for image content validation. The binary image content validation is performed with the help of a deep learning model named MobileNetV3-Large. The skin images are forwarded to the deep learning model named EfficientNet-V2M for multi-class dermatological disease classification into seven different disease categories. The experimental results of the binary skin and non-skin image classification model achieve an accuracy of 98.6%, while the seven-class dermatological disease classification model achieves an accuracy of 90%. The results of the proposed model are promising for tele-dermatology services.

Keywords—component, dermatological disease detection, deep learning, tele-dermatology, image quality assessment, MobileNetV3-Large, EfficientNet-V2M.

I. INTRODUCTION

Dermatological diseases are one of the prominent health issues affecting the global population. People from various ages and different geographic locations are impacted by these diseases. Therefore, it becomes imperative to diagnose these diseases at the earliest and in the most accurate manner to prevent the advancement of these diseases. With the rise of telemedicine practices, tele-dermatology has come to play an important role in addressing geographically isolated and under-resourced environments. The majority of tele-dermatology practices use images obtained from patients using their own devices like smartphones and the Internet. Figure 1 shows some of the examples of dermatological images obtained under real-world conditions.

Despite the advantages offered by tele-dermatology practices, these systems face considerable challenges with regard to the quality and content of the obtained images. The obtained images from the patient may not always be clear and may be noisy and poorly illuminated. The images

may also be distorted and partially occluded. Apart from these issues, irrelevant images may also be obtained from the patient. All these issues may affect the accuracy of the predictions obtained from the automated systems for the detection of dermatological diseases.

The current study aims at the task of dermatological image classification using deep learning techniques, specifically convolutional neural networks (CNN) that learn complex lesion/texture features. Although existing studies mainly emphasize multi-class classification tasks assuming that the input images are clinically relevant and of sufficient quality, such an assumption is not always valid in tele-dermatology practice. To address this drawback of existing methods, we propose a multi-stage deep learning framework that includes the evaluation of image integrity and the removal of blurred images using the excision module. Then, a binary classification of skin/non-skin images using the MobileNetV3 Large model is performed. Finally, the remaining skin images are classified into one of the seven classes of dermatological diseases using the EfficientNetV2M model. The experimental results show that the accuracy of the binary classification task is 98.6%, while the overall accuracy of the seven-class classification task is 90%.



Figure 1 Disease Sample

II. RELATED WORKS

Significant progress has been made in the field of automated dermatological disease detection using deep learning approaches. A major breakthrough was reported by Esteva et al. on the application of convolutional neural networks for the classification of skin cancers, which

exhibited dermatologist-level performance with a high area under the curve (AUC) value, similar to that of human dermatologists [1]. This research marked the beginning of the application of deep neural networks for dermatological image analysis. Subsequent research on the application of transfer learning with public datasets ISIC and HAM10000 reported a classification accuracy of more than 85%, with competitive sensitivity and specificity under controlled experimental conditions [5], [13]. This research confirmed that deep neural networks can be used for the effective classification of dermatological images.

Systematic reviews and meta-analyses have also been conducted on the diagnostic potential of artificial intelligence in dermatology. A comprehensive systematic review reported that convolutional neural network-based approaches frequently exhibit high accuracy, precision, recall, and F1 score on benchmark datasets [3], [4]. A broader evaluation of artificial intelligence approaches for dermatology reported that many approaches exhibit a high AUC value, while sensitivity and specificity vary across heterogeneous populations [6], [7].

Recent research has also emphasized issues with regard to data set heterogeneity and diversity in terms of the population under investigation. Studies that have attempted to evaluate the generalization performance of the model over a wide range of populations have reported fluctuations in terms of accuracy and AUC while applying the model over a different data distribution [11], [19]. Moreover, meta-analytical evaluation of different data sets has emphasized that the validation protocols need to be sensitive to issues of sensitivity, specificity, and false-positive rates for the successful application of the model [6]. This has also emphasized that a high classification accuracy is not the only key determinant for successful application in the context of a telemedicine system.

While previous research has mainly focused on optimizing multi-class classification metrics with regard to architectural refinements or ensemble learning strategies [8], [18], less emphasis has been given to the development of a pre-classification validation framework. Few frameworks have attempted to incorporate image integrity, blur, or binary relevance filtering before moving on to the final stages of disease classification. This is because, in the context of a tele-dermatology system, images that are noisy, blurred, or non-dermatological in nature are common, which can be addressed by the incorporation of progressive stages of validation.

In this context, the present work has attempted to extend the state of the art by integrating a framework of image validation, binary filtering, and multi-class classification for the development of a more precise system for tele-dermatology.

III. METHODOLOGY

The framework that is proposed in this paper utilizes a structured multi-stage deep learning approach that is primarily focused on improving the robustness and reliability of the system in real-world tele-dermatology settings. Contrary to the conventional end-to-end classification-based approaches, the proposed methodology incorporates image validation, relevance filtering, and classification in a sequential manner to tackle the problems

of low-quality images. The system architecture is shown in Figure 2.

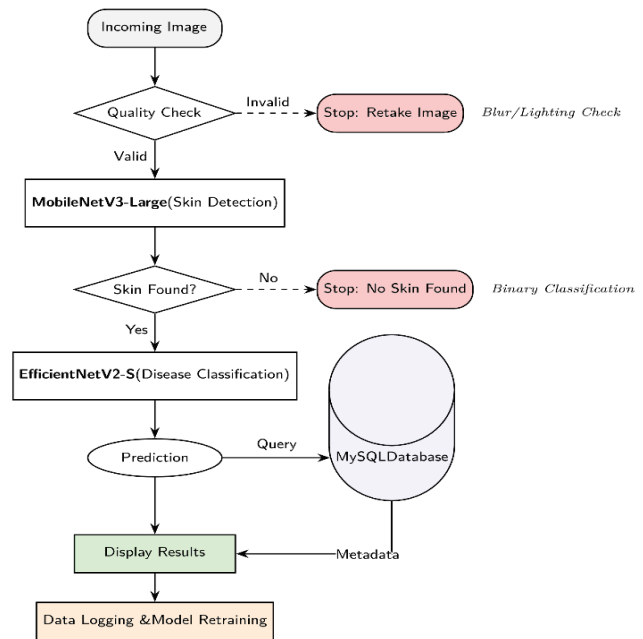


Figure 2 Project Architecture

A. Incoming Image Acquisition

The system processes the input images, which are obtained through the use of smartphones or other web-based devices within tele-dermatology applications. Considering the fact that the images are captured in uncontrolled conditions, the input images could be blurred, noisy, or suffer from improper lighting, or could be compressed or contain unwanted information. Therefore, the preprocessing and validation steps are critical before the actual disease classification.

B. Layer 1: Image Integrity and Validity Assessment

The first stage is responsible for the assessment of the integrity of the images, as well as the blur, to ensure that the images are diagnostically viable before they are classified. Tele-dermatology images are prone to degradation or blurring since they are taken under uncontrolled conditions. The integrity of the images is checked by assessing the readability of the files, where corrupted images are rejected.

Blur detection is then performed using the variance of the Laplacian method. For a grayscale image $I(x, y)$, the Laplacian operator is defined as:

$$\nabla^2 I(x, y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \quad (1)$$

The sharpness score is computed as the variance of the Laplacian response:

$$\sigma_L^2 = \text{Var}(\nabla^2 I) \quad (2)$$

If the variance σ^2/L is below a predefined threshold T , the image is considered blurred and rejected; otherwise, it proceeds to the next stage. This step prevents low-detail images from degrading classification performance.

C. Binary Skin–Non-Skin Classification

Consequently, images identified as entering the validity stage are then passed to a binary classification model based on the MobileNetV3-Large architecture. This stage adjudicates the images to ascertain whether the image bears dermatological skin content. The binary classification helps to eliminate irrelevant images or those that may have been submitted in error. In the event the image does not bear a classification of skin, the process ends.

Images that pass the validation stage are forwarded to a binary classifier based on the MobileNetV3-Large architecture. Let the input image be denoted as x . The binary classifier estimates the probability:

$$P(x), y \in \{0, 1\} \quad (3)$$

where $y = 1$ represents skin and $y = 0$ represents non-skin. The final decision is obtained as:

$$\hat{y} = \underset{y \in \{0,1\}}{\operatorname{argmax}} P(x) \quad (4)$$

If $\hat{y} = 0$, the image is rejected and processing is terminated. Experimental evaluation shows that this stage achieves 98.6% accuracy, demonstrating effective elimination of irrelevant inputs prior to disease inference.

D. Multi-Class Disease Classification

The validated skin images are then processed by a deep convolutional neural network, which is based on the EfficientNet-V2M architecture. This classifier is used to classify the diseases into the seven classes. EfficientNet-V2M was chosen for its accuracy and computational efficiency. The classifier produces softmax outputs corresponding to the classes, and the class with the highest softmax output is chosen as the final diagnosis. The classifier is able to obtain 90% accuracy over the seven classes.

E. Database Lookup and Output Logging

After the prediction, the system performs a database lookup to obtain the information related to the disease corresponding to the predicted class. The final output displays the identified dermatological condition to the user. Additionally, the input image and prediction logs are saved for further evaluation, which can be used to improve the system.

IV. CHALLENGES

Despite the promising results obtained by the proposed framework of multiple stages of deep learning, a number of issues remain to be considered with regard to the implementation of the proposed system in real-world tele-dermatology applications. Variability in image acquisition conditions, dataset variability, constraints with regard to the generalization of the proposed system, and operational constraints remain to be considered.

A. Data Collection and Image Variability

The collection of dermatology images, which are diverse and representative, is another major challenge. Images acquired through tele-dermatology are taken in uncontrolled conditions, where the quality of the images is compromised due to the variety of smartphone cameras used to capture the images. The variability in the quality of the images, including resolution, lighting, focus, and background, may reduce the generalization capability of the classifier. Ensuring the diversity of the classes, along with maintaining the distribution across the classes, is another major challenge.

B. Image Quality Degradation

The quality of the images received in tele-dermatology is compromised, showing blur, noise, or occlusion. Even though the blur is detected using the Laplacian-based blur detector, establishing the threshold is not trivial, especially considering the variety of conditions. If the threshold is set too high, the classifier may not be able to generalize across all the conditions, whereas if the threshold is set low, the classifier may be influenced by the low-quality images, thereby affecting the accuracy.

C. Binary Filtering Misclassification Risk

The binary filtering of the skin image from non-skin images holds a crucial position in the system. Even though the system has a high accuracy of 98.6%, misclassifications of the images might cause the system to miss valid images in the disease classifier. Therefore, the sensitivity of the system at the binary filtering level should be high.

D. Generalization to Real-World Deployment

The models might face difficulties when deployed in a real-world environment. The models might experience a reduction in performance when deployed in a real-world environment. The performance of the models might decrease when deployed in a real environment. The models might face difficulties in generalizing the performance of the models in a real-world environment.

E. Computational Efficiency and Deployment Constraints

Though the EfficientNet-V2M has a high efficiency in terms of its performance compared to its computational efficiency, there are certain constraints in the deployment of the system. The system might face difficulties in the deployment of the system in a tele-dermatology system.

V. EXPERIMENTAL RESULTS

A. Experimental Setup

The experimental evaluation used a specially prepared dermatological image dataset that was compiled from a variety of publicly available sources. For the proposed framework, a dataset was prepared that corresponds to a two-stage framework: one for binary classification (skin or non-skin) and another for seven-class disease classification.

For the binary classification stage between skin and non-skin images, 14,000 images were used in total: 7,000 dermatological skin images and 7,000 non-skin images. The

dataset was prepared in a balanced manner in order to avoid training bias. The distribution of this dataset is provided in Table I.

TABLE I. BINARY MODEL DATASET DISCRPTION

Name	No. of Images
skin	7000
Non-skin	7000

In the multi-class classification of dermatological diseases, around 42,000 images were used, and these were equally distributed among the seven dermatological classes, with each class containing 6,000 images. This ensures that the problem of class imbalance is reduced, and the model converges properly. The description of the multi-class dataset is given in Table II.

TABLE II. 7-CLASS MODEL DATASET DISCRPTION

Name	No. of Images
Bacterial Infection	6000
Psoriasis	6000
Eczema	6000
Fungal Infection	6000
Normal/Healthy skin	6000
Parasitic infection	6000
Viral infection	6000

Binary Classification Model Configuration

The binary classifier was created by employing a MobileNetV3-Large model with pretrained weights from ImageNet. The last classification layer was discarded by setting include_top to False. The model was frozen initially to leverage the pretrained feature representation. Next, a global average pooling layer was used, and a dropout layer was added with a dropout rate of 0.3 to prevent overfitting. The final output layer consists of a single neuron with a sigmoid activation function.

The training protocol and hyperparameters are as follows:

- Loss function: Binary cross-entropy
- Optimizer: Adam
- Metrics: Accuracy, AUC, Precision, Recall
- Learning rate: = 0.01
- Fine-tuning learning rate: 1×10^{-5}
- Fine-tuning strategy: Last 100 layers unfrozen (excluding BatchNormalization layers)
- Early stopping and ReduceLROnPlateau were applied to prevent overfitting

Seven Class Disease Classification Model Configuration

The classifier for the seven classes of diseases was implemented utilizing EfficientNetV2-M, predicated on the ImageNet-pretrained weights, along with global average pooling. The initial freezing of the backbone network occurred during the Stage 1 training. The dropout layer,

with the dropout rate set to 0.35, was utilized prior to the final fully connected layer, which utilized the softmax activation function over the seven classes of diseases.

The model was trained using:

- Loss function: Sparse categorical cross-entropy
- Optimizer: Adam
- Metrics: Accuracy, Top-3 Accuracy, Top-5 Accuracy
- Fine-tuning strategy: Last 60 layers unfrozen
- Fine-tuning learning rate: 1×10^{-5}
- Early stopping and learning rate scheduling were employed

During initial training, the model demonstrated progressive convergence, with validation performance improving across epochs before fine-tuning.

B. Evaluation Metri

The proposed framework was evaluated by utilizing the classification metrics: Accuracy, Area Under the Receiver Operating Characteristic Curve (AUC), Precision, Recall, Top-3 Accuracy, and Top-5 Accuracy. These metrics provide a comprehensive evaluation of the binary and multi-class classification performance.

For binary classification, let TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

The overall accuracy is defined as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

Precision measures the proportion of correctly predicted positive samples:

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

Recall (Sensitivity) measures the proportion of actual positives correctly identified:

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

The Area Under the ROC Curve (AUC) evaluates the model's ability to distinguish between classes across varying decision thresholds.

For multi-class classification with K classes, the predicted probability for class i is obtained using the softmax function:

$$P(x) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (8)$$

Top-k accuracy evaluates whether the correct class appears within the top K predicted probabilities. Formally,

$$Top - k = \frac{1}{N} \sum_{n=1}^N \mathbf{1}(y_n \in \hat{Y}_n^{(k)}) \quad (9)$$

where represents the set of top K predicted classes for sample n , and $1(\cdot)$ is the indicator function.

The accuracy trajectory of the binary classification model during the training process is shown in Figure 3. The figure shows a steady improvement in the performance of the model. This shows the effectiveness of the features learned by the model. The trajectory of the validation accuracy in the same figure shows a similar trend to the training accuracy. This shows the robust performance of the model in terms of generalization.

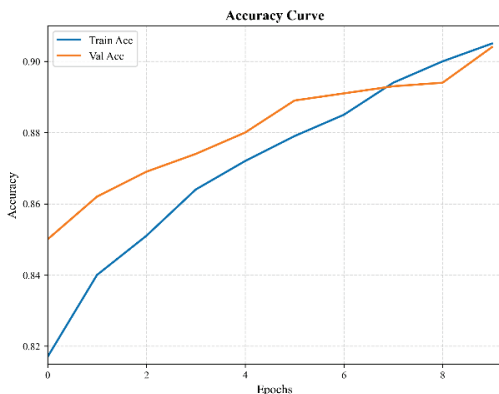


Figure 3 training and validation accuracy

Figure 4 shows the training and validation loss curves for the binary classification model. The figure shows a steady decrease in the loss of the model. This shows the robust performance of the model in terms of convergence. The fluctuations in the validation loss of the model in the later stages do not show any signs of overfitting.

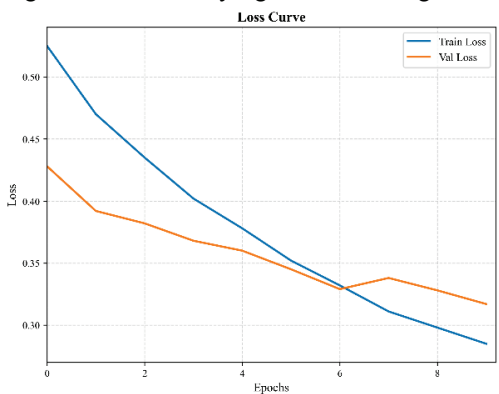


Figure 4 training and validation loss

Figure 5 shows the final evaluation metrics of the binary classifier. The figure shows high values of the evaluation metrics of the binary classifier. The high values of the metrics show the robust performance of the binary classifier in terms of discriminative ability. The binary classifier can effectively differentiate between skin and non-skin images.

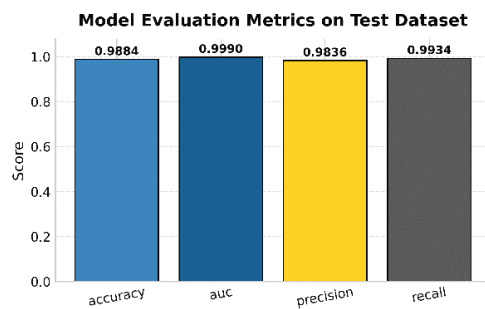


Figure: Comparison of key evaluation metrics — Accuracy, AUC, Precision, and Recall.

Figure 5 Binary classification model

C. Comparative Performance Analysis

Table III shows a comparative analysis of the proposed framework in comparison to other contemporary state-of-the-art approaches. In the binary classification problem for the MobileNetV3-Large classifier, it is observed that it achieves 98.88% accuracy, thereby ensuring a high degree of discriminative capability for eliminating non-dermatological images. In the seven-class disease classification problem, it is observed that the EfficientNetV2-M achieves 90.0% accuracy in comparison to other contemporary approaches in [5], [10], and [15], thereby ensuring a high degree of performance for the proposed multi-stage architecture.

TABLE III
 COMPARISON WITH STATE-OF-THE-ART METHODS

Author	Model	Task Type	Classification Accuracy (%)
Esteva et al. [1]	Inception-v3 CNN	Binary (Skin Cancer)	91.0
Ahammed [5]	CNN (ISIC + HAM10000)	Multi-class	88.0
Aldhyani [10]	Lightweight CNN	Multi-class	89.7
Shetty et al. [15]	Deep CNN	Multi-class	87.4
Proposed (Binary)	MobileNetV3-Large	Binary (Skin vs Non-Skin)	98.88
Proposed (7-Class)	EfficientNetV2-M	Multi-class (7 classes)	90.0

The experiments have been conducted in a high-performance GPU environment by utilizing NVIDIA L40s accelerators. The training was performed by utilizing a setup that includes four GPUs, each providing 48 GB memory and a peak compute rate of 362 TFLOPS, as well as 48 CPU cores. This enabled the effective training and fine-tuning of both architectures, namely MobileNetV3-Large and EfficientNetV2-M. The chosen hardware provided a balanced trade-off between computation and utilization to ensure stable convergence for large-scale dermatological images.

Beyond achieving strong quantitative performance, the proposed multi-stage framework demonstrates practical capability in recognizing dermatological conditions under real-world imaging scenarios. As illustrated in Figure 6, the EfficientNetV2-M model predicts the presented clinical image as eczema. The comparatively moderate confidence score reflects the visual similarity that often exists between inflammatory skin conditions, where overlapping lesion patterns and diffuse erythema can introduce classification ambiguity. Such cases highlight the inherent complexity of multi-class dermatological diagnosis, particularly when images are captured under non-controlled conditions. Despite this challenge, the model successfully identifies clinically relevant texture and distribution patterns associated with eczema. This qualitative example emphasizes the importance of robust feature extraction and balanced multi-class training in improving diagnostic reliability across heterogeneous clinical images.

Prediction: Eczema
Confidence: 0.2021

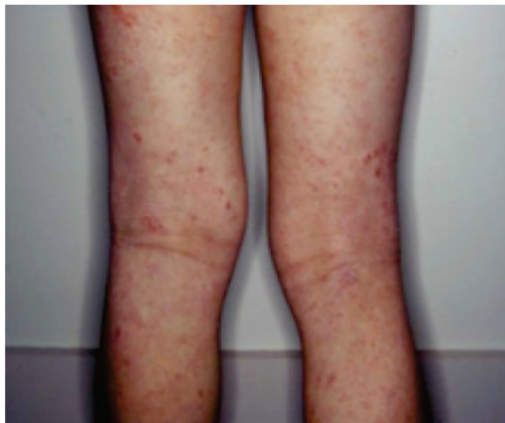


Figure 6 Seven-class prediction result (Eczema)

VI. CONCLUSION

The present study proposes a structured, multi-stage deep learning framework for the automated detection of dermatological diseases in tele-dermatology settings. Unlike conventional classifiers, which directly perform multi-class classification, the proposed framework is composed of separate stages for integrity verification, Laplacian-based blur detection, binary skin-non-skin classification using the MobileNetV3-Large model, and seven-class disease detection utilizing the EfficientNetV2-M model. These stages are integrated to ensure the reliability of the proposed framework, which prevents corrupted, low-quality, or irrelevant data from affecting the final classification results.

The experimental results show the binary classifier stage to be highly accurate, with 98.88% accuracy, high discriminative power, and the proposed EfficientNetV2-M model to attain 90% accuracy on the balanced, multi-source dermatology dataset. The convergence curves show the proposed model to be stable, with minimal overfitting. The qualitative results also validate the proposed model's potential to detect clinically significant patterns in real-world scenarios.

The proposed framework is significant in the context of tele-dermatology, where the structured, multi-stage approach to deep learning is likely to improve the reliability of the proposed framework, thereby enhancing its applicability in real-world scenarios. The proposed framework is also computationally efficient, thereby making it suitable for the automated detection of dermatological diseases.

REFERENCES

- [1] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–118, 2017. doi: 10.1038/nature21056.
- [2] K. A. Muhaba, A. B. Rashid and M. Anwar, "Automatic skin disease diagnosis using deep learning from clinical images," *JMIR Dermatol.*, vol. 3, no. 1, e18438, 2020. doi: 10.2196/18438.
- [3] S. Choy, A. Tschandl, E. G. Codella et al., "Systematic review of deep learning image analysis for skin diseases," *npj Digital Medicine*, vol. 6, p. 18, 2023. doi: 10.1038/s41746-023-00914-8.
- [4] H. Jeong, N. Lee, J. W. Kim and K. Y. Kim, "Deep learning in dermatology: A systematic review," *Frontiers in Medicine*, vol. 9, 2022. doi: 10.3389/fmed.2022.9841357.
- [5] M. Ahammed, "A machine learning approach for skin disease detection using ISIC 2019 and HAM10000 datasets," *Comput. Biol. Med.*, vol. 145, p. 105515, 2022. doi: 10.1016/j.combiomed.2022.105515.
- [6] Z. R. Cai, D. A. Ghaznavi, J. H. C. and C. J. Walsh, "Assessing the performance of artificial intelligence models in clinical diagnosis: A meta-analysis for skin disease severity," *Br. J. Dermatol.*, vol. 193, no. 5, pp. 847–859, 2025. doi: 10.1093/bjd/ljad123.
- [7] M. Salinas et al., "Skin cancer detection using deep learning—a review," *JAMA Dermatol.*, 2023. doi: 10.1001/jamadermatol.2023.1025.
- [8] H. Liu, J. Yang, Y. Song and P. Xia, "A skin disease classification model based on multi-scale channel attention," *Sci. Rep.*, vol. 15, 2025. doi: 10.1038/s41598-025-90418-0.
- [9] I. Abunadi, A. A. Khan and M. Mirjat, "Deep learning and machine learning techniques of skin lesion classification," *Electronics*, vol. 10, no. 24, p. 3158, 2021. doi: 10.3390/electronics10243158.
- [10] H. H. Aldhyani, "Multi-class skin lesion classification using a lightweight CNN approach," *Sensors*, vol. 22, no. 9, p. 2048, 2022. doi: 10.3390/s22092048.
- [11] A. Aquil, S. Lee and H. Kim, "Early detection of skin diseases across diverse populations using deep learning," *Inf. Sci.*, vol. 16, no. 2, p. 152, 2025. doi: 10.3390/info16020152.
- [12] Q. Wu, H. Yin, H. Chen and M. Sun, "A deep learning, image-based approach for automated diagnosis of inflammatory skin diseases," *Ann. Transl. Med.*, vol. 8, no. 9, p. 581, 2020. doi: 10.21037/atm.2020.04.39.
- [13] R. Rashmi and G. R. Ravi, "Automatic skin disease diagnosis using pre-trained MobileNet-V2 model," *Comput. Biol. Med.*, vol. 142, p. 105452, 2022. doi: 10.1016/j.combiomed.2022.105452.
- [14] A. Rasheed et al., "Automatic eczema classification in clinical images based on hybrid deep neural network," *Comput. Biol. Med.*, vol. 153, p. 105807, 2022. doi: 10.1016/j.combiomed.2022.105807.
- [15] B. Shetty et al., "Skin lesion classification of dermoscopic images using deep learning," *Sci. Rep.*, vol. 12, 2022. doi: 10.1038/s41598-022-22644-9.
- [16] S. Abbas et al., "Intelligent skin disease prediction using transfer learning," *Sci. Rep.*, vol. 15, 2025. doi: 10.1038/s41598-024-83966-4.
- [17] O. Attallah, "Explainable deep learning classification of skin cancer using feature fusion," *Inf. Fusion*, vol. 92, p. 102881, 2024. doi: 10.1016/j.inffus.2024.102881.
- [18] S. M. Thwin and et al., "Skin lesion classification using deep ensemble model," *Appl. Sci.*, vol. 14, no. 13, p. 5599, 2024. doi: 10.3390/app14135599.
- [19] J. Vieira et al., "Deep learning approaches for skin lesion detection and classification," *Electronics*, vol. 14, no. 14, p. 2785, 2025. doi: 10.3390/electronics14142785.
- [20] M. Fiaz et al., "Explainable hybrid deep learning framework for precise skin lesion segmentation and classification," *Front. Med.*, vol. 12, p. 1681542, 2025. doi: 10.3389/fmed.2025.1681542.