



A LIGHTWEIGHT ATTENTION-GUIDED MULTI-SCALE ENSEMBLE FRAMEWORK FOR ROBUST SKIN DISEASE CLASSIFICATION

Vignesh M¹, Rajeswari K²

¹PG Student, Department of Computer Science and Engineering, Mahendra Engineering College, Namakkal, India

²Assistant Professor, Department of Computer Science and Engineering, Mahendra Engineering College, Namakkal, India

¹<http://orcid.org/0009-0007-3736-6822>, ²<http://orcid.org/0009-0005-8047-1420>

Email: vignesh.vigneshkumar1@outlook.com, rajeswarik@mahendra.info

ARTICLE INFO

Article History

Received: March 30, 2026

Reviewed: August 3, 2026

Accepted: September 3, 2026

Published: September 30, 2026

Keywords:

Skin Disease Diagnosis,
Deep Learning Framework,
Dermoscopic Image Analysis,
Noise Reduction,
Multi-Scale Feature Extraction,
Attention-Based Feature Fusion,
Lightweight Ensemble Learning,
Image Preprocessing,
Computationally Efficient
Classification

ABSTRACT

Early treatment and minimization of the impact of the dangerous complications, such as skin cancer, is impossible without adequate diagnosis of skin disease. Nevertheless, the commonly used traditional diagnostic methods rely on the visual inspection of dermatologists, and thus the procedure is subjective, time-consuming, and unavailable in remote and resource-limited environments. To address these drawbacks, the following paper presents a small and multi-stage deep learning architecture that can be used to classify intelligent skin diseases with a balance between significant accuracy and high computing efficiency. The given system is expected to be efficient in terms of processing the dermoscopic images and is still applicable to being implemented on the devices with limited computer capabilities. The framework starts with the Adaptive Skin Image Noise Reduction Module (ASINRM) that improves image quality by using illumination normalization, noise reduction, contrast enhancement and intensity standardization. The step of preprocessing minimizes variability brought by the lighting conditions and imaging artifacts. The processed images will then be fed to the Multi-Scale Residual Feature Extraction Network (MSRFEN) to undergo accurate lesion segmentation to remove the affected area and isolate it with the rest of the healthy skin. This step of segmentation enhances localization of features and makes the later learning to be centered on clinically relevant areas. MSRFEN also simultaneously derives discriminative lesion detail at several spatial scales, including both fine-grained texture and coarse detail structural features, required to classify to high accuracy. The next is the Cross-Scale Adaptive Feature Fusion Network (CSAFFN) which combines the multi-resolution features via an attention-guided fusion mechanism which focuses on the most informative representations. To classify, the Lightweight Ensemble Skin Disease Classification Module (LESDCM) employs several effective learners which work in parallel to produce solid probability predictions. Lastly, the Skin Lesion Robust Ensemble Prediction Aggregation (REPASL) module is an ensemble-based prediction aggregation module that is adaptive and uses confidence-based weighting and consensus boosting techniques to deliver a dependable final diagnosis. It is experimentally proven that the suggested framework has high classification accuracy, high generalization performance, and low computational complexity, which makes it appropriate to be applied in the real-life clinical and mobile healthcare setting.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Diagnosis of skin diseases is a cornerstone activity in the field of dermatology which entails the identification, division, and categorization of a typical skin lesions in order to allow timely therapy and eliminate serious complications like melanoma and other chronic skin conditions. Over the last few years, artificial intelligence has grown at an incredibly fast rate, and automated diagnostic systems that use deep learning have become effective in helping clinicians to analyse dermoscopic images. Overall, the automated skin disease classification working process involves image acquisition, illumination correction and noise suppression preprocessing, lesion segmentation, discriminative features extraction, feature fusion, and final classification with the help of the deep neural networks.

The improved localization of boundary and the ability to detect lesion regions has been shown in recent segmentation schemes that use network feature aggregation module and edge-enhanced attention mechanism [1]. Strong classification systems with ensembles

of distillation networks of deep neural networks have further improved predictability and made them available in the mobile diagnostic systems [2]. Serial attention mechanisms have been proposed to achieve a more and more refined segmentation of lesions by better capturing the contextual and spatial dependencies [3]. Moreover, the networks of deep attention branches have enhanced classification with greater sensitivity of informative lesion features [4]. Multi-scale feature map fusion techniques have also been useful in extracting fine-grained texture details and global structure patterns thus enhancing the accuracy of skin cancer classification [5]. Although all these developments have been made, a number of limitations do exist. A great number of current architectures are computationally intensive and need high-end hardware resources, which restricts their use in real-time and resource-constrained situations. Segmentation and classification performance may be adversely affected by sensitivity to changes in illumination, imaging artifacts and background noise. Moreover, intricate attention systems and rich ensemble designs can expand the model size and the inference time, and affect scalability in mobile healthcare use. To overcome these limitations, the current paper offers a lightweight multi-stage deep learning model, which would combine high diagnostic accuracy with the ability to be performed with low computational cost. The suggested architecture incorporates the adaptive preprocessing, multi-scale residual features extraction, attention-based features fusion, and ensemble-based classification using robust prediction aggregation. The framework is expected to offer reliable, scalable, and clinically applicable skin disease diagnosis to fit real-world and mobile healthcare systems by optimizing the representation of the features and the efficiency of the model.

1.1 THE CONTRIBUTIONS OF THIS WORK ARE THE FOLLOWING:

- To introduce an Adaptive Skin Image Noise Reduction Module (ASINRM) that improves the quality of dermoscopic images through illumination normalization, noise reduction, contrast intensification, and the standardization of intensities, hence making it robust to a wide range of imaging conditions.
- To create a Multi-Scale Residual Feature Extraction Network (MSRFEN) that can extract small scale texture features and global structure at once and thus can effectively model the complex lesion features.
- To present a Cross-Scale Adaptive Feature Fusion Network (CSAFFN) which incorporates complementary multi-resolution features through attention-based fusion system to produce a compact and discriminative feature representation.
- To develop Lightweight Ensemble Skin Disease Classification Module (LESDCM) that makes use of parallel efficient learners to improve predictive stability to improve generalization capability and decrease the computational complexity.
- To make it fit in a real-time clinical setting and be implemented in a mobile healthcare setup, by making the whole framework computationally efficient and scalable.

The paper has been organized in the following way: Section 1 of the paper presents an introduction to automated skin disease diagnosis and the need of lightweight multi-stage deep learning structures. Section 2 presents the literature review related to the current study, which includes research on the dermoscopic image segmentation, attentional mechanisms, multi-scale feature extraction plans and the ensemble-based classification methods. Section 3 describes the framework that is proposed in details and justifies the functional roles of ASINRM, MSRFEN, CSAFFN, LESDCM, and REPASL modules. Section 4 introduces the experimental set up, performance evaluation metrics, comparative analysis, and robustness evaluation of the experimental set up on various imaging conditions. Finally, a conclusion of the paper is provided in Section 5, which summarizes the main findings, limitations and future research directions.

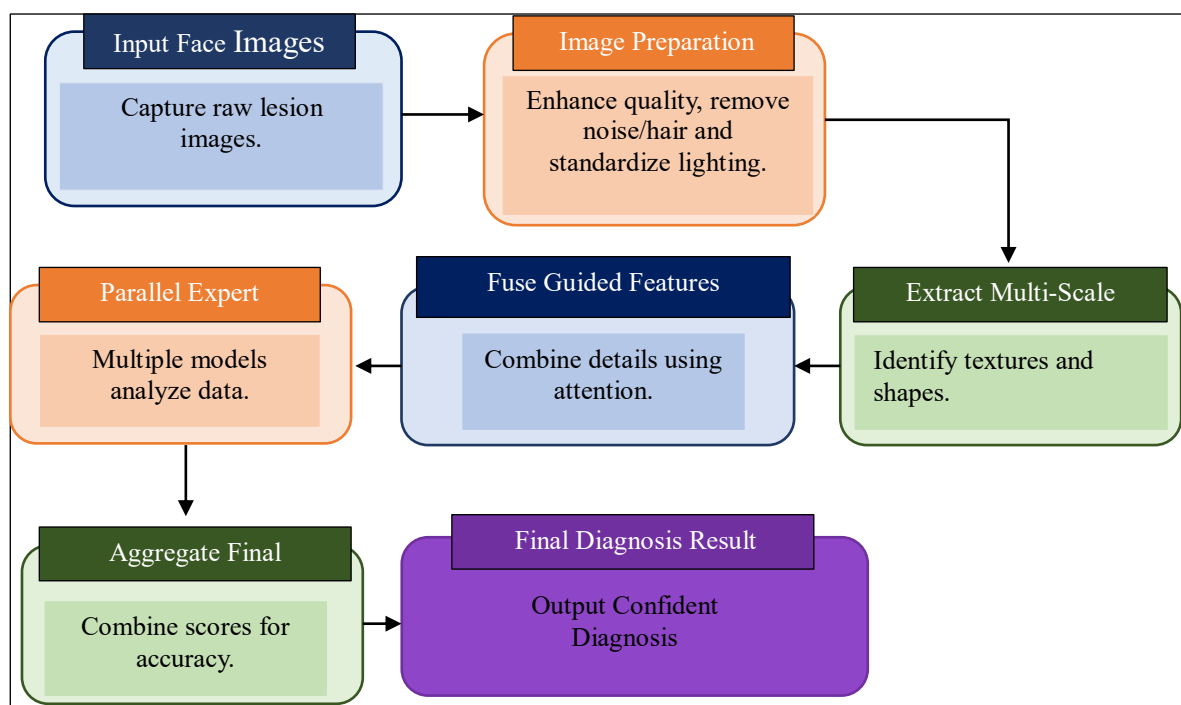


Figure 1: Basic Architecture of the Proposed Skin Disease Classification Framework.

Source: Authors, (2026).

Figure 1 represents a simplified deep learning structure to automate the process of classifying skin diseases. The first stage is raw lesion image capture after which the image preparation phase follows where the image is optimized by removing noise and standardizing illumination. Multi-scale feature extraction is subsequently done to obtain needed details on texture and structural patterns

of the lesion areas. An attention-guided mechanism is used to fuse the extracted features in order to produce a complete and discriminative representation. A series of parallel expert models are used to analyse the fused features in order to enhance robustness and predictive power. Lastly, these combined predictions are summed up to deal with a sure and trustworthy diagnostic output, which can be used in clinical decision support.

II. LITERATURE SURVEY

The recent developments in automated analysis of skin lesions have focused on multi-scale learning, attention, and integration of multi-modes in enhancing the diagnostic accuracy. A better method of multi-scale feature fusion to skin lesion segmentation was proposed in [6], where multi-scale feature aggregation strategy has been designed which is applied to integrate hierarchical representations at different resolutions. The suggested approach improved the definition of the boundaries of lesions through the inclusion of additional significant semantic information as compared to shallow spatial information. However, the architecture had low contrast lesion area vulnerability and had a higher computational complexity because of the repetitiveness of the feature fusion processes. Multi-Scale Attention Fusion with Lesion-Area Focus (MSAF-LAF), which proposed knowledge-enhanced dermoscopic skin lesion classification framework, was proposed in [7]. In order to integrate multi-scale contextual information and highlight the areas of clinical importance, the lesion-area attention modules were added to this strategy. It improved classification accuracy, however, as it utilized more complicated attention fusion blocks, inference time was higher and not as applicable to lightweight deployment. A Deep Multi-Modal Skin-Imaging-Based Information-Switching Network (DMSI-ISN) was introduced in [8]. This network dynamically chose the complementary features to be used in recognition of skin lesions with various imaging modalities and adaptive information-switching process.

This multimodal integration improved the stability in various imaging scenarios but the multiple imaging inputs also reduced its use in typical dermoscopic data and further complicated the process of data acquisition. A lightweight and explainable model called HIMViT (Hybrid Interpretable Modified Vision Transformer) was proposed in [9] according to a modified architecture of Mobile Vision Transformer to attain efficient classification of skin diseases. The model was better in terms of computational efficiency and it had interpretability mechanisms but the limited parameter design limited deep contextual learning in very complex lesion patterns. Moreover, a framework called Attention-Guided Learning with Feature Reconstruction (AGL-FR) was offered in [10] to diagnose the skin lesion using a combination of both clinical and ultrasound images. The technique used attention guided feature reconstruction to improve learning cross-modal representations. Although it was shown to have a better diagnostic performance, multimodal imaging and reconstruction overhead added complexity into the architecture and computation. In general, the accuracy of segmentation and classification is greatly increased in cases of current research, but the aspects of computational efficiency, scalability, and the ability to adhere to a range of imaging conditions are still open research questions

Table 1: Literature Survey Analysis of Existing Skin Lesion Classification Methods.

Author / Year	Type of Algorithm	Context Extraction	Fragmentation Process	Similarity Computation Approach
Behara et al., 2024 [11]	Hybrid Active Contour + Attention-Guided Capsule Network	Attention-based capsule feature learning	Active contour snake model for lesion segmentation	Capsule agreement and softmax classification
Thwin & Park, 2024 [12]	Deep CNN Ensemble Model	Hierarchical deep feature extraction	No explicit segmentation	Soft voting and ensemble averaging
Erbay et al., 2025 [13]	CNN-Transformer Ensemble	CNN local + Transformer global context modeling	Implicit convolution-based feature maps	Weighted ensemble fusion
Chiu et al., 2025 [14]	Two-Stage Voting Ensemble	Deep CNN feature extraction	Dermoscopic lesion-focused preprocessing	Majority voting strategy
Kumar Lilhore et al., 2024 [15]	Hybrid U-Net + Improved MobileNet-V3	Encoder-decoder multi-scale features	U-Net based lesion segmentation	Optimized MobileNet classification
Liu et al., 2021 [16]	Multiscale CNN Ensemble	Multi-scale convolutional context learning	No explicit segmentation	Ensemble probability averaging
Hammad et al., 2025 [17]	HybridFusionNet with Explainability	Hybrid feature fusion with attention maps	Implicit feature refinement	Confidence-based hybrid classification
Manzoor et al., 2025 [18]	Dual-Stage Segmentation-Classification	Context-aware deep features	Explicit lesion segmentation	Sequential probability mapping
Gallazzi et al., 2025 [19]	Sequential Segmentation & Classification	CNN contextual learning	Sequential lesion segmentation	Supervised similarity learning
Fiaz et al., 2025 [20]	Explainable Hybrid Deep Framework	Attention-guided representation learning	Segmentation-based mask generation	Hybrid probability-based classification

Source: Authors, (2026).

Table 1 includes the comparative analysis of the recent segmentation and classification methods of skin lesion according to the type of algorithm, strategy of extracting the context, process of fragmentation, and similarity computation method. The majority of the studies utilize hybrid deep learning models which are convolutional neural networks, transformers, capsule networks, or ensemble approaches to improve the representation of features and classification stability. Some of these approaches include explicit lesion

segmentation with the use of either U-Net or active contour models to enhance localization of regions prior to classification, whereas others use implicit feature-level learning without explicit segmentation in between. Multi-scale convolution or attention mechanisms or global transformer-based modeling are the major ways to extract context and capture both local texture and overall structural data. Similarity computation and final decision-making are normally done by ensemble voting, probability averaging and confidence-weighted fusion. On the whole, both of these approaches prove to be more accurate in their diagnostic methods, but most of them have complicated structures that add computational load and reduce the efficiency with which they can be deployed in real-time. Recent advancements in automated skin disease analysis have focused on integrating multi-scale feature learning, attention mechanisms, and explainable architectures to improve classification accuracy and lesion localization.

To classify and localize multiclass skin lesions, a new Network-Level Fused Deep Architecture with Explainable Artificial Intelligence (NLFDA-XAI) was suggested in [21]. This model has the characteristics of several convolutional layers to obtain both local and global lesion data and explainable AI capabilities to offer clear decision-making. Although this method had better interpretability and accuracy in localization, it was computationally expensive and thus might not be able to run in real-time in situations with limited hardware. A Cross-Scale Attention and Multi-Layer Feature Fusion YOLOv8 (CSAMFF-YOLOv8) architecture was proposed in [22] to detect the target of skin diseases. This approach builds on the cross-scale attention modules to combine the features of several layers of YOLOv8 and emphasize the improvement of lesion-specific areas. Though the CSAMFF-YOLOv8 was able to localize and detect lesions accurately, the use of the YOLO-based architecture added the inference time that restricted the effectiveness of high-throughput lesion screening. A model named Attention-Based MobileNet-V2 (AB-MN2V) was created in [23] to conduct a more advanced classification of skin diseases through a combination of multi-scale feature extraction with lightweight attention modules. This solution is efficient to deploy on a mobile device and maintain accuracy.

Nevertheless, it is not good enough to deal with very complicated lesions because its design is lightweight making it be limited to deep feature representation. Similarly, a Multi-Residual Attention Network (MRAN) was created in [24] that aims to capture hierarchical lesion representations which comprise numerous residual attention blocks. MRAN enhanced feature discrimination as well as classification robustness but introduced much architectural complexity which augmented both training time and computational cost. In [25], a systematic review identified the recent deep learning-based dermatological condition detection techniques, with the focus on the fact that the multi-scale learning and attention mechanisms enhance the effectiveness; but the problems are related to the heterogeneity of the datasets, variability of the annotation, and real-time implementation into clinical practice. The paper [26] involved the use of traditional Machine Learning Approaches (ML-ASND) to diagnose skin neoplasm using classical classifiers and hand-crafted features to identify lesions. Though these approaches are computationally light and interpretable, they tend to perform poorly than deep learning approaches because they represent limited features. A Multi-Scale Combined Efficient Channel Attention Module (MS-CECAM) was also introduced in [27] in order to extract multi-scale lesion features and use channel attention to emphasize salient areas.

This was a better way to classify but it came with more model parameters and increased training. Enhanced Skin Lesion Segmentation and Classification Ensemble Model (ESLSC-EM) in [28] also showed better robustness on the combination of a number of deep learners but the complexity of the models and the cost of inference was high. In [29], it was proposed to use a Multiclassification Model Using Inception-v2 (MM-I2V) to recognize skin diseases by using Inception-v2 multi-scale feature maps to enhance discriminative capabilities. Although very accurate, the model was very large in terms of memory and computation power. Lastly, the Deep Learning Framework of Early Diagnosis and Classification of Skin Cancer Lesions (DLF-EDCSCL) in [30] merged convolutional layers (with adaptive feature extraction) with early malignant lesion Point detection. Although it is useful in detection at an initial stage, the model is computationally expensive and cannot perform optimally without large annotated datasets. Taken together, these works demonstrate the significant advancements in skin lesions classification, however, they suggest that the accuracy, interpretation, and computational efficiency can be traded at the cost of developing lightweight, explainable, and scalable frameworks suitable to clinical usage.

III. PROPOSED METHODOLOGY

The methodology begins with Adaptive Skin Image Noise Reduction Module (ASINRM), which is charged with the preprocessing of skin images so that they become better and consistent throughout the dataset. This module is used to correct illumination variations to compensate the difference in lighting and eliminates sensor noise, noise in the environment as well as standardizing the skin texture to give consistent images that can be used to extract features. Since it cleans the input images and corrects inconsistencies, ASINRM guarantees that the next steps are based on meaningful lesion patterns and not irrelevant artifacts, and this is very important to enhance the accuracy and the strength of the system. After preprocessing, Multi-Scale Residual Feature Extraction Network (MSRFEN) is used to extract meaningful features of the images. In MSRFEN, a mix of residual connections and multi-scale convolutions is applied in order to extract the fine-grained and broad-level details. Detailed features like colour variations, texture, and micro-patterns are spared whereas larger structures like shape of lesion, irregularity of its borders, and total area are also captured. The multi-level feature extraction promotes the system to distinguish among skin conditions that can resemble visually, and approximate how such lesions characteristics can be understood. After the extraction of the features, the Cross-Scale Adaptive Feature Fusion Network (CSAFFN) incorporates the inter-scale information. Through the mechanism of attention, CSAFFN also focuses on the most informative information and rejects redundant or irrelevant data. The effect of this is a consistent and detailed description of every skin lesion to enhance the quality of the inputs into the classification phase. The Lightweight Ensemble Skin Disease Classification Module (LESDCM) is then used to classify the lesions. In this module, numerous lightweight models are used and work simultaneously, each one bearing on the various subsets of extracted features. LESDCM alleviates overfitting by integrating the predictions of these models and enhances the generalisation of the system to changes in skin types, lesion size and image quality. The ensemble approach enables the system to use different attitude towards the data to classify the data more precisely.

Lastly, the Robust Ensemble Prediction Aggregation Skin Lesions (REPL) module summarizes the results of LESDCM. REPASL will give different predictions an adaptive weight, depending on the confidence score, which will combine the strengths of both models. The stage yields very final predictions which are reliable and accurate enough to be strong even in very difficult cases. The general procedure is effective, flexible, and applicable to deploying on a device with limited computing capabilities, and it is high-performing in an extensive range of skin lesion types.

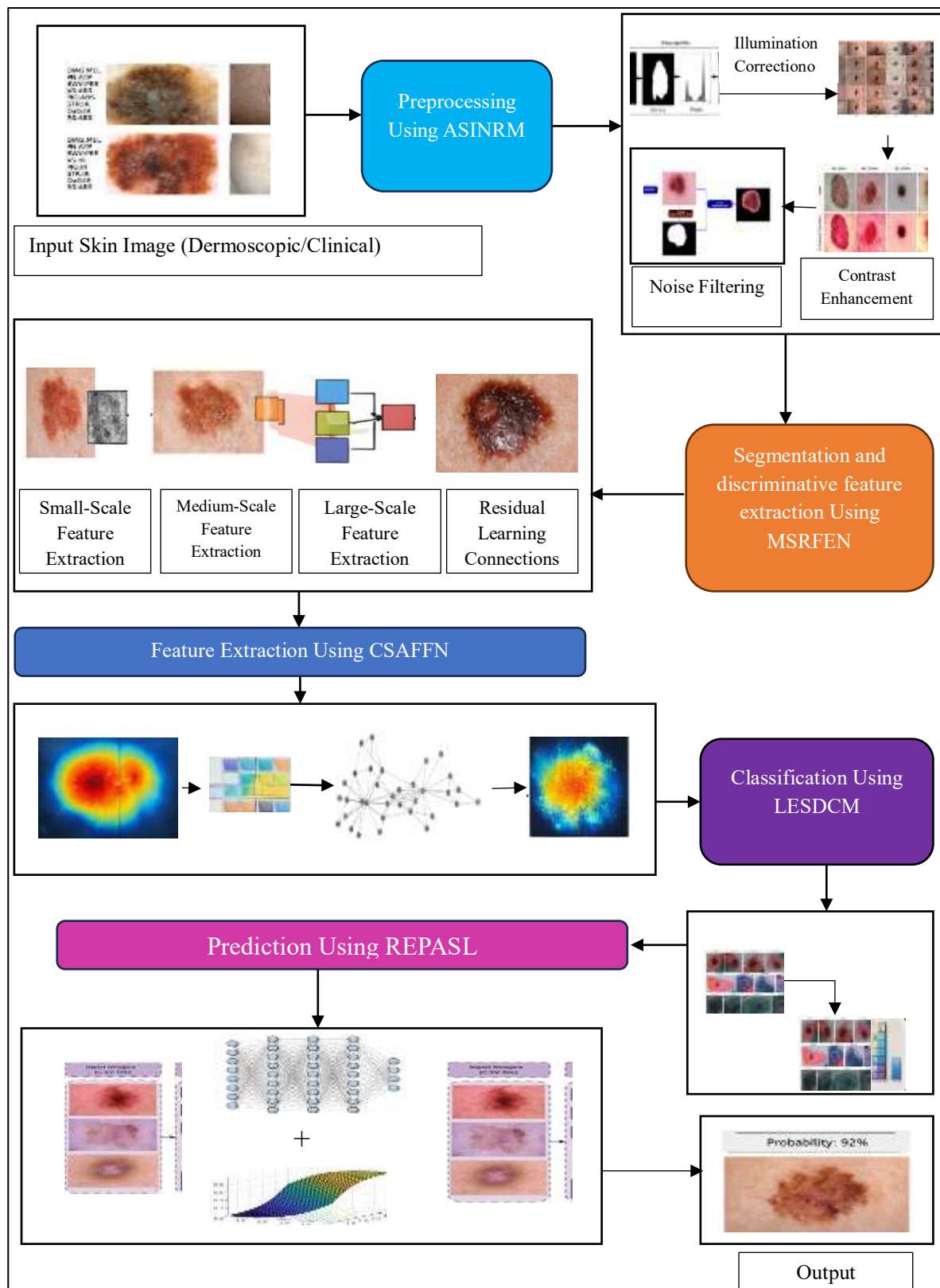


Figure 2: Proposed Lightweight Multi-Stage Deep Learning Architecture for Skin Disease Classification.

Source: Authors, (2026).

Figure 2 The proposed architecture The proposed architecture is a lightweight and computationally-efficient multi-stage deep learning framework that classifies skin diseases automatically. It firstly involves preprocessing with the Adaptive Skin Image Noise Reduction Module (ASINRM) to correct illumination, filter noise and enhance contrast to enhance the quality of dermoscopic images.

These refined images are then used by Multi-Scale Residual Feature Extraction Network (MSRFEN) which extracts the lesion characteristics at various spatial resolution using residual learning mechanisms. This facilitates the acquisition of fine-grained texture as well as the large-scale structural pattern which is vital in the correct diagnosis. The Cross-Scale Adaptive Feature Fusion Network (CSAFFN) is then used to combine multi-level features by using an attention-based fusion approach to create a powerful feature representation. The composite features are then sent to the Lightweight Ensemble Skin Disease Classification Module (LESDCM) which uses several effective learners to come up with probabilistic forecasts. Lastly, The Robust Ensemble Prediction Aggregation Skin Lesions (AB) (REPASL) module is the one that takes the outputs of the ensemble, weighting them by confidence to obtain a highly trustworthy

final diagnosis. In general, the architecture guarantees high accuracy, good generalization ability, and low computational complexity, which makes it appropriate in the real-world clinical and mobile healthcare application.

III.1 DATASET DESCRIPTION

The dataset for skin diseases used in this project is a multi-class classified dataset of dermoscopic images that contains ten separate diagnostic groups. The dataset contains labelled clinical images that were collected, curated, and organised into their own respective class folders for the purpose of supervised learning (i.e. images that are classes/classified in terms of diagnoses). The ten classes of diagnostic images are: Eczema (1,677 images), Melanoma (15,750 images), Atopic Dermatitis (1,250 images), Basal Cell Carcinoma (3,323 images), Melanocytic Nevus (7,970 images), Benign Keratosis-type Lesions (2,624 images), Psoriasis and Lichen Planus-like Disorders, Seborrheic Keratosis and other types of Benign Tumours, Tinea Ringworm-type Candidiasis and Fungal Infection-type Infections, and Warts and Viral-type Infections (Molluscum and Viral-type Infections). The dataset contains significant class imbalance because both Melanoma and Melanocytic Nevus have a number of images in contrast to both fungal-related disease and inflammatory related diseases (e.g. Eczematous dermatosis). There is considerable variation in the images regarding resolution of the images, as well as variations regarding the conditions under which the images were taken, variations with respect to the skin tones of the individuals in the image, variations regarding the size of the lesions, variations with respect to the complexity of the lesions, and variations with respect to background artifacts from the images; therefore, these images represent real-world variability within medicine. Each of the images contains the appropriate disease label for use in multi-class classification and segmentation-based analysis. The dataset contains the data that is used to develop deep learning-based models for segmenting lesions and extracting features from lesions, as well as accurately classifying skin diseases. The overall variability of the types of pathologic patterns across the ten total classes of skin diseases provides the proposed framework with robustness and will enhance the framework's ability to generalise to perform well in real-world clinical moraine diagnostic systems.

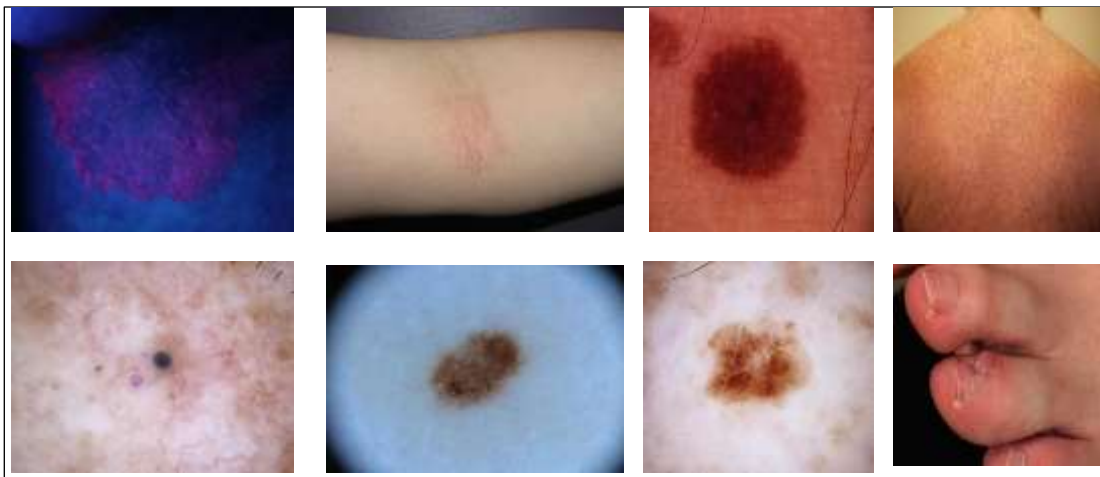


Figure 3: Sample Dermoscopic and Clinical Images from the Skin Disease Dataset.

Source: Authors, (2026).

Figure 3 shows examples of skin diseases from the dataset that have been used in this project. The examples of skin diseases show variations in the colour of the lesions, size of the lesions, texture of the lesions, irregularity of the borders of the lesions and location of the lesions that exist across various skin diseases. The skin disease samples that were used in this project contain both dermoscopic images as well as non-dermoscopic (standard clinic) images. Therefore, the examples show many different pigmentation patterns, vascular components, scaling of the lesions, and surface topography of the lesions.

III.2 ADAPTIVE SKIN IMAGE NOISE REDUCTION MODULE (ASINRM)

The Adaptive Skin Image Noise Reduction Module (ASINRM) is the first and initial block in the proposed skin disease classification system, whose function is to ensure that the final input images are of high quality and can be analyzed further. Images of skin used to classify skin images are usually diverse in inconsistencies such as differences in illuminations, artifacts in the ground, sensor noise, and that of the skin texture among people. Such ambiguities may negatively impact deep learning model performance because noisy or non-standardized images can result in poor performance in extracting features and poor classification performance. ASINRM deals with these issues through the use of an adaptive preprocessing technique that improves the clarity of images, balances colour and brightness, and eliminates unnecessary noise. The module starts with a step of carrying out illumination correction, which balances the lighting of each image to an equal level, so that the variation in brightness or shadowing will not deceive the process of the feature extraction. After this, ASINRM implements noise reduction algorithms to eliminate random sensor noise as well as background artifacts such as tiny spots, hair or other non-lesion components that may affect analysis. It is an adaptive step which implies that the amount of correction and filtering is altered dynamically with regards to the quality and the characteristics of each image.

The module also normalizes the skin texture and colour value distribution and brings all images into a comparable range of pixel intensities and clears small variations without modifying specific features of lesions. ASINRM creates clean and normalized images, which makes the input data provided to the Multi-Scale Residual Feature Extraction Network (MSRFEN) a representation of the underlying skin lesions. This improves feature extraction network capabilities to extract finer detail of the lesion edges, textures, pigmentation patterns and structural irregularities whilst retaining the overall context of the lesion. Moreover, the system is adaptive and

can be used to work with different image quality and this feature contributes to its resistance to fluctuations in light, skin color, and other conditions of image capture.

$$I_{\text{illum}} = I_{\text{raw}} \times \frac{L_{\text{mean}}}{L_{\text{local}}} \quad (1)$$

In Equation 1, I_{raw} is an original input image of the skin dataset. L_{local} refers to the brightness of a pixel area, and L_{mean} refers to the average of the illumination of the image. The fixed image I_{illum} modifies local variations in brightness to produce a uniform image, and shadows or irregular lighting do not pose a problem to further feature detection.

$$N_{\text{est}} = \sigma(I_{\text{illum}}) + \lambda_{\text{sensor}} \quad (2)$$

To estimate the noise N_{est} in the illumination-corrected image, Equation 2 is used. In this case $\sigma(I_{\text{illum}})$ is the standard deviation of the pixel intensities, which is a measure of random variations in brightness, and λ_{sensor} takes into consideration natural sensor noise. This step measures the amount of undesired variation to be decreased prior to extraction of features.

$$I_{\text{denoise}} = I_{\text{illum}} - \alpha \cdot N_{\text{est}} \quad (3)$$

Then the image that has undergone denoising is denoted by I_{denoise} in Equation 3 and is represented as I_{denoise} = illumination corrected image, which is computed by subtracting the weighted noise estimate. The alpha value balances the suppression of the noise and makes sure that important lesion details are not lost and superfluous noise is minimized.

$$T_{\text{norm}} = \frac{I_{\text{denoise}} - \mu_{\text{patch}}}{\sigma_{\text{patch}}} \quad (4)$$

In equation 4, the texture of the denoised image is normalized, where μ_{patch} and σ_{patch} , mean and standard deviation respectively, of pixel intensities in local patches of the picture. T_{norm} is used to guarantee that the local variations in contrast are normalized and that the model is concentrated on the actual patterns of the lesion and not the local variations in lighting or texture.

$$C_{\text{std}} = \frac{T_{\text{norm},c} - \mu_c}{\sigma_c} \quad (5)$$

In equation 5, individual color channels of the normalized image are standardized. $T_{\text{norm},c}$ is the normalized value of a particular color channel c , while μ_c and σ_c are the mean and the standard deviation of the channel across the data. This process provides uniformity in color representation and this is essential in the identification of lesions patterns based on color.

$$I_{\text{prep}} = f_{\text{clip}}(C_{\text{std}}) \quad (6)$$

The end result of equation 6 is the final preprocessed image I_{prep} which is now prepared to extract features. The f_{clip} is the function that makes sure that all the pixel values are in the valid ranges of the model input. This last output combines the outputs of illumination correction, noise reduction, texture normalization and color standardization giving a clean and consistent image to the downstream modules such as feature extraction and classification. ASINRM is a preprocessing step used in skin image to perform skin illumination correction, noise cleaning and texture standardization. It will also keep all the images consistent and of high quality so that further modules can concentrate on the meaningful lesion patterns. It can remove irrelevant artifacts in the dataset by adapting to changes in light and the skin type. The module enhances the effectiveness and strength of feature extraction. On the whole, ASINRM provides a framework to base on the classification of skin diseases.

III.3 MULTI-SCALE RESIDUAL FEATURE EXTRACTION NETWORK (MSRFEN)

As a result of the preprocessing phase, the Multi-Scale Residual Feature Extraction Network (MSRFEN) is the central unit in the case of lesion segmentation and discriminative features extraction of skin images that are enhanced. Once illumination correction, noise suppression, texture normalization, and color standardization have been performed in ASINRM, the refined images are sent to MSRFEN where lesion localization and representation learning are done accurately. First, MSRFEN conducts a lesion segmentation to separate the affected area against the rest of the healthy skin tissue hence guaranteeing that the future learning of features in the target area only concentrates on parts of the body that are clinically relevant. This step of segmentation minimizes the interference of the background and enhances spatial attention to pathological areas. After segmentation, both fine-grained and global-level features that are crucial to making correct classifications will be extracted by the network. The convolutional layers with different receptive fields are introduced in the architecture, which is able to capture small-scale structure variability, including pigmentation variations, micro-textures and subtle edge irregularities, and at the same time learn large-scale structural variability, including lesion symmetry, shape, and border distortion. The multi-scale representation makes the network able to distinguish similar lesions to the eye (with minor structural and textural variations) based on a multi-scale representation. Residual connections within MSRFEN preserve feature integrity by allowing information to bypass intermediate layers, preventing degradation of low-level details and mitigating vanishing gradient issues in deeper layers. Operationally, the segmented lesion region is processed at multiple scales to generate complementary feature maps that emphasize distinct lesion characteristics. These maps are integrated through element-wise addition, effectively combining local and global contextual information into a unified representation. Activation and normalization layers further stabilize feature distributions, enhancing generalization across variations in lighting conditions, skin tones, and lesion appearances.

Ultimately, MSRFEN produces robust multi-scale segmented feature representations that capture both subtle and prominent lesion attributes. These refined features are subsequently transmitted to the Cross-Scale Adaptive Feature Fusion Network (CSAFFN) for further prioritization and enhancement, ensuring highly discriminative inputs for final classification.

$$F_{\text{small}} = I_{\text{prep}} * K_{\text{small}} \quad (7)$$

In Equation 7, I_{prep} is the preprocessed skin image obtained from ASINRM and K_{small} is a convolution kernel with a small receptive field. Before fine-scale feature extraction, the segmented lesion region is emphasized so that convolution focuses primarily on the isolated pathological area rather than surrounding healthy skin. The small feature map F_{small} captures fine-grained lesion characteristics such as subtle texture variations, pigmentation irregularities, and micro-structural boundaries within the segmented region. This localized segmentation-guided extraction enhances sensitivity to early-stage and minor lesion abnormalities.

$$F_{\text{medium}} = I_{\text{prep}} * K_{\text{medium}} \quad (8)$$

In Equation 8, the preprocessed image I_{prep} is convolved with a medium-sized kernel K_{medium} to produce the feature map F_{medium} . At this stage, segmentation ensures that intermediate-level structural patterns are learned specifically from the delineated lesion mask, minimizing background interference. This map captures regional morphological properties such as localized shape distortions, color transitions, and boundary irregularities, strengthening lesion-focused representation.

$$F_{\text{large}} = I_{\text{prep}} * K_{\text{large}} \quad (9)$$

In Equation 9, K_{large} is a convolution kernel with a large receptive field applied to I_{prep} , producing F_{large} . Segmentation-guided processing allows the large-scale receptive field to model the global spatial organization of the isolated lesion, including symmetry, border sharpness, and overall geometry. These global contextual descriptors are essential for distinguishing visually similar disease categories.

$$F_{\text{residual}} = F_{\text{small}} + F_{\text{medium}} + F_{\text{large}} + I_{\text{prep}} \quad (10)$$

The multi-scale features are integrated through element-wise addition in Equation 10 along with the original preprocessed image I_{prep} as a residual connection. Because segmentation has already refined lesion localization, the residual fusion preserves both segmented spatial attention and hierarchical contextual information. The resulting feature map F_{residual} retains micro-level and macro-level lesion descriptors without losing structural integrity.

$$F_{\text{norm}} = \text{ReLU}(\text{BatchNorm}(F_{\text{residual}})) \quad (11)$$

In Equation 11, Batch Normalization stabilizes and scales the residual features F_{residual} , followed by the ReLU activation function to introduce non-linearity. Normalization further enhances segmentation-aware representations by suppressing irrelevant background activations and reinforcing lesion-specific discriminative responses. The output F_{norm} therefore emphasizes clinically meaningful patterns.

$$F_{\text{MS}} = \text{Concat}(F_{\text{norm}}, F_{\text{small}}, F_{\text{medium}}, F_{\text{large}}) \quad (12)$$

In Equation 12, the normalized residual features F_{norm} are concatenated with the original multi-scale feature maps to produce the final multi-scale representation F_{MS} . Since all feature maps are derived from segmentation-focused processing, the concatenated representation comprehensively captures lesion-specific micro-textures, intermediate morphology, and global structural attributes. This ensures that the Cross-Scale Adaptive Feature Fusion Network (CSAFFN) receives a refined and spatially consistent lesion representation. MSRFEN therefore performs segmentation-guided multi-scale feature extraction, preserving localized lesion boundaries while enhancing discriminative learning. By integrating segmentation with residual multi-scale learning, the network achieves improved robustness against background noise and visual variability. The resulting feature set provides a rich and clinically relevant representation for adaptive fusion and final classification.

III.4 CROSS-SCALE ADAPTIVE FEATURE FUSION NETWORK (CSAFFN)

The Multi-Scale Residual Feature Extraction Network (MSRFEN) is employed to obtain rich multi-scale features of the skin lesion, but Cross-Scale Adaptive Feature Fusion Network (CSAFFN) is important to merge the features and create a consistent and well-informed representation of the lesion. The multi-scale feature maps, which represent the fine details, intermediate patterns, and the global structure information, are complementary in nature but can have redundant or less informative elements in the proposed system. To deal with this, CSAFFN uses an adaptive fusion scheme that focuses on the most appropriate features amongst scales and repress irrelevant or redundant features. This will make sure that the network is attentive to the most important aspects of lesion morphology, texture, and colour patterns which are critical in the classification. The fusion process commences by considering the multi-scale feature maps F_{MS} created by MSRFEN or attention based weighting of scales. These attention weights are dynamically learned on the dataset and this enables the network to give more significant attention to those features that are more discriminatory to differentiate between classes of skin diseases. As an illustration, in an image where fine-grained texture can be vital, the attention system will place more importance on small scale information but in one with a large lesion shape or boundary one of the large scale features will be given greater emphasis. This adaptation mechanism makes sure that the fused representation is made to suit the properties of each of the images in the set. After the attention-based weighting, CSAFFN integrates scales in the form of elements-based addition and concatenation, which integrate all the scales into one unified feature map. Such integration retains the fine and coarse details as well as preserving interactions of features at various scales, which increases the representational power of the network. Also, normalization and non-linear activation functions are used to ensure stability and enhance the discriminability of the fused features. The CSAFFN output is high-dimensional feature representation that has been refined and contains local and global lesion data and is ready to undergo to the next stage of classification.

CSAFFN enhances the capability of the network to be more adaptive by consolidating multi-scale information to generalize across several different images within the dataset and across lesions size, color, shape, and skin type differences. All in all, CSAFFN facilitates the system to be not only precise and strong enough but also able to display the most significant patterns that can be used in the real-life situation to classify the skin disease reliably.

$$F_{\text{input}} = F_{\text{MS}} \quad (13)$$

In the equation 13, F_{input} is the multi-scale feature maps produced by MSRFEN. Such maps consist of small, medium and large-scale characteristics, which depict the texture, color and shape of the lesions on the skin. F_{input} is the initial stage of the fusion process and it has an abundant pool of complementary features that will be optimised and incorporated in the further equations.

$$A_{\text{small}}, A_{\text{medium}}, A_{\text{large}} = \text{Softmax}(W_{\text{attn}} \cdot F_{\text{input}} + b_{\text{attn}}) \quad (14)$$

Attention weights in each scale of feature are computed in equation 14. W_{attn} and b_{attn} are parameters of the attention module which are learnable. The values of the weights $A_{\text{small}}, A_{\text{medium}}, A_{\text{large}}$ are such that they add up to one across scales. This helps the network to adjust to features depending on their relevance to classifying each lesion in the dataset.

$$F_{\text{weighted}} = A_{\text{small}} \cdot F_{\text{small}} + A_{\text{medium}} \cdot F_{\text{medium}} + A_{\text{large}} \cdot F_{\text{large}} \quad (15)$$

In equation 15, the weights of attention are multiplied with the respective feature map of MSRFEN. F_{weighted} The feature representation is scaled-adapted, i.e., the network puts more weight on the most discriminative features of the dataset. As an example, the texture can be fine and dominate in some types of lesions whereas global shape can dominate over it in others.

$$F_{\text{fusion}} = F_{\text{weighted}} + F_{\text{input}} \quad (16)$$

In Equation 16, a weighted version of the feature map F_{weighted} is added to the original multi-scale input F_{input} element-wise. This type of residual-like integration makes sure that no original information is damaged during this combination of adaptive features, and both raw and contextually accentuating features are preserved to provide a more desirable representation of lesions.

$$F_{\text{norm}} = \text{BatchNorm}(F_{\text{fusion}}) \quad (17)$$

Equation 17 is used to implement batch normalization on the fused feature map F_{fusion} . F_{norm} is used to normalize the distribution of the features, minimizes internal covariate shift and stabilizes the learning process. This is a step that enhances generalization in the various images in the dataset even the changes in the size, color and the type of skin.

$$F_{\text{CSAFFN}} = \text{ReLU}(F_{\text{norm}}) \quad (18)$$

Equation 18 produces the final output of the Cross-Scale Adaptive Feature Fusion Network, F_{CSAFFN} , is the result of applying a ReLU to the normalized fused features in equation 18 as given below. This last composite image brings out the most informative characteristics in all the scales rendering this image prepared to join the classification module. F_{CSAFFN} is strong, discriminative, and biased towards the capturing of subtle and global lesion patterns in the dataset. CSAFFN combines cross-scale and attention-based weighted multi-scale features. It highlights the most informative aspects and inhibits the redundant information. The network generates a high-dimensional representation of every lesion that is coherent. CSAFFN enhances feature discriminability in different skin types by changing image features based on the features of a particular image. It trains powerful attributes to the ensemble classification module.

III.5 LIGHTWEIGHT ENSEMBLE SKIN DISEASE CLASSIFICATION MODULE (LESDCM)

After the Cross-Scale Adaptive Feature Fusion Network (CSAFFN) produces the refined multi-scale feature representation, these features are passed to the Lightweight Ensemble Skin Disease Classification Module (LESDCM) for final lesion prediction. LESDCM is also meant to utilize the strength of having many lightweight neural models at a time instead of using one big classifier. The fused features CSAFFN have been used in this proposed system as the input into one of the parallel branches, which address various aspects or subsets of the features, which may include texture, color or structural patterns. This guarantees that no one direction could prevail over the decision making mechanism and the ensemble will be in a better position to detect minute variations among the skin lesion classes that exist in the dataset. The structure of all branches in LESDCM is designed to be computationally efficient, and yet has adequate capacity to learn discriminative patterns. Through the distribution of the learning on various models, the module will minimize the chance of overfitting, especially with even datasets with different types of lesions and different image quality. The ensemble generates preliminary predictions by each branch and gives a compilation of probability vectors that approximate the probability of the lesion being in each category.

Such predictions represent the complementary nature of the knowledge acquired by various branches that allows the module to integrate the various viewpoints in the analysis of the same input image. To further improve strength, the ensemble incorporates confidence-based weighting, in which branches that are more suitable about certain classes are more weighted. This adaptive weighting will make sure that greater signals in the dataset are boosted whereas weak and less certain predictions are deemphasized. Consequently, the output of LESDCM is a combination of the information of several specialized classifiers, which gives the initial but a dependable prediction of every input image. It is handed over to the next Robust Ensemble Prediction Aggregation for Skin Lesions (REPASL) module that further consolidates and refines prediction on the branches to generate the ultimate skin disease. Overall, LESDCM is an important component of the suggested system as it offers a balance, adaptive, and computationally efficient classification mechanism, which could accommodate the variability and complexity of the images of skin lesions in the dataset and predict with high accuracy the various types of skin lesions.

$$F_{\text{branch}} = F_{\text{CSAFFN}} \quad (19)$$

F_{branch} is the input of CSAFFN in the Equation 19. These are the multi-scale features which are fused with fine-grained texture and world structural information of the skin images. F_{branch} is used as the input of all the ensemble branches, where each lightweight classifier is provided with a discriminative representation of every lesion.

$$P_{b_i} = \text{Softmax}(W_{b_i} \cdot F_{\text{branch}} + b_{b_i}) \quad (20)$$

Equation 20 is used to calculate prediction of the i th branch of the ensemble. W_{b_i} and b_{b_i} are learnable weights and biases of that branch. The output is transformed into class probabilities P_{b_i} of each of the skin disease categories in the dataset by the Softmax function. Every branch has a distinct area of specializing in detecting a specific number of features, and this enables a wide range of perspectives in the same input image.

$$C_{b_i} = \max(P_{b_i}) \quad (21)$$

The confidence score C_{b_i} of each branch is computed using equation 21 as the maximum probability in each branch prediction vector. The score shows the confidence of the branch on its best-predicted class, which is used in proportioning the significance of the prediction of each branch in the final ensemble.

$$P_{b_i}^{\text{weighted}} = C_{b_i} \cdot P_{b_i} \quad (22)$$

The probability of each branch P_{b_i} in Equation 22 is multiplied by its confidence score C_{b_i} to create a weighted probability vector $P_{b_i}^{\text{weighted}}$. This means that the more certain the branches, the larger the influence and this comes in handy especially when dealing with variations in the appearance of the lesion in the dataset.

$$P_{\text{ensemble}} = \sum_{i=1}^N P_{b_i}^{\text{weighted}} \quad (23)$$

The weighted sum of all of the N ensemble branches is used to construct the initial ensemble prediction P_{ensemble} in equation 23. The weighted sum of the probabilities adds the system together on an array of views of the lesion, making it more robust and able to generalize across a variety of different skin types and lesion types.

$$P_{\text{LESDCM}} = \text{Normalize}(P_{\text{ensemble}}) \quad (24)$$

The aggregation of the probabilities P_{ensemble} of any given ensemble is normalized in equation 24 to make sure they add up to 1 resulting in P of any given ensemble P_{LESDCM} , the final product of the LESDCM module. This probability vector is the distribution of likelihood of the lesion to be a member of each of the classes in the dataset, which is inputted into the next REPASL module to make final predictions. LESDCM uses a set of light models to provide initial estimates to every image. Both branches have subsets of features of interest, which minimizes overfitting and maximizes generalization. The use of confidence-based weighting makes sure that predictions that are reliable take over the ensemble. LESDCM is a combination of several models, which generates a powerful probability vector. It produces corrective and flexible classification results to be aggregated.

III.6 ROBUST ENSEMBLE PREDICTION AGGREGATION FOR SKIN LESIONS (REPASL)

After the Lightweight Ensemble Skin Disease Classification Module (LESDCM) has made initial projections on each skin image, the Robust Ensemble Prediction Aggregation of Skin Lesions (REPASL) module additionally amalgamates and checks them to form the final, very high-confidence classification. LESDCM gives a list of probability vectors of each image, indicating the forecasts of numerous lightweight branches. Although LESDCM already takes confidence-weighted outputs into account, the difference in the appearance of the lesion, the lighting, and the skin type of the dataset may cause uncertainty in certain predictions. REPASL handles this by doing a strong aggregation which is adaptive with regards to the confidence of each branch as well as the consistency of the prediction of the ensemble. This is done by first gathering together all probability vectors $P_{\text{"LESDCM"}}$ of the ensemble branches. REPASL uses an adaptive deposit strategy in which prediction with a high score on the score of confidence is provisional and those with a low score have a low influence. This weighting is necessary so that strong signals in the dataset, such as unusual texture, color, or shape patterns, are exaggerated and that ambiguous features do not affect the final decision so much. REPASL also checks the consistency of the branches in addition to weighting and focuses on classes that are agreed upon by different branches lowering the chances of misclassification due to outliers or noisy predictions.

After the weighting and summation, the module normalizes the obtained final output to obtain a probability distribution across all classes of skin diseases. This last probability vector is the best prediction of the input image that involves the detailed multi-scale features that are obtained by MSRFEN and CSAFFN as well as the collective intelligence of LESDCM. The design of REPASL makes the system resistant to changes in the dataset such as changes in lesion size, color, and border irregularity, which makes it apply to the real-world scenarios where the image quality and skin type can vary significantly. In general, REPASL is the decision-making unit of the suggested system, which will help convert several initial projections into one and one certain and accurate answer. The module is capable of high reliability, accuracy and generalization due to the combination of adaptive weighting, ensemble consensus, and normalization, which completes the entire end-to-end workflow of image preprocessing to multi-scale feature extraction, adaptive fusion, ensemble classification and final prediction. This renders the system able to provide reliable classification of skin diseases using a wide range of datasets and clinical settings.

$$P_{\text{input}} = P_{\text{LESDCM}} \quad (25)$$

P_{input} in Equation 25 is the initial probability vectors generated by LESDCM on every input image. In each of the vectors, the probable occurrence of all skin disease categories according to the ensemble branches is predicted. This input is the starting point of REPASL which will start to refine and compile these predictions into one strong output of each image in the dataset.

$$C_{\text{branch}} = \max(P_{\text{input}}) \quad (26)$$

Equation 26 computes the score of confidence C_{branch} of each branch in the ensemble by picking the largest probability in P_{input} . This score indicates the confidence of the branch in its most likely class and will be used to weigh its contribution during the final aggregation whereby more accurate predictions will have a bigger impact.

$$P_{\text{weighted}} = C_{\text{branch}} \cdot P_{\text{input}} \quad (27)$$

Equation 27 is a multiplication of the prediction vector of each branch with their respective confidence score to generate P_{weighted} . The adaptive weighting amplifies the impact of high-confidence predictions in the dataset and reduces less confident outputs so that the strength of signals in the skin lesion images prevails in the ultimate decision.

$$P_{\text{aggregate}} = \sum_{i=1}^N P_{\text{weighted},i} \quad (28)$$

The weighted prediction of each of the N ensemble branches is added in Equation 28 to come up with the aggregated prediction vector, $P_{\text{aggregate}}$. This is done by combining several views of each skin image which incorporates fine, intermediate, and gross features of lesions to come up with a strong initial prediction.

$$P_{\text{consensus}} = P_{\text{aggregate}} \cdot \text{AgreementFactor}(P_{\text{weighted}}) \quad (29)$$

Equation 29 brings in the consistency or agreement factor across the branches which enhances the predictions made by the different branches that are consistent. $P_{\text{consensus}}$ reinforces classes with high levels of ensemble agreement, increasing the contribution of data respondents, which is imperative to deal with variability in the appearance of skin lesions in the data set.

$$P_{\text{REPASL}} = \text{Normalize}(P_{\text{consensus}}) \quad (30)$$

To bring the consensus vector $P_{\text{consensus}}$ to a normalized form and make sure that all probabilities add up to 1, equation 30 is used to normalize this consensus vector. This last probability vector is the strong and highly predictive prediction of the skin lesion type, which is fit to be reported or deployed. It incorporates multi scale characteristics, adaptive fusion, ensemble classification and confidence weighting, which give it good generalization across the heterogeneous set of data. REPASL takes results of LESDCM and gives the end result of the skin disease prediction. The adaptive weighting and consensus mechanisms are based on high-confidence and predictive consistency. This module transforms the final vector to obtain a strong and reliable probability distribution. It improves the system accuracy and reliability to variability of datasets. REPASL provides reliable classification that is fit to be used in the real world.

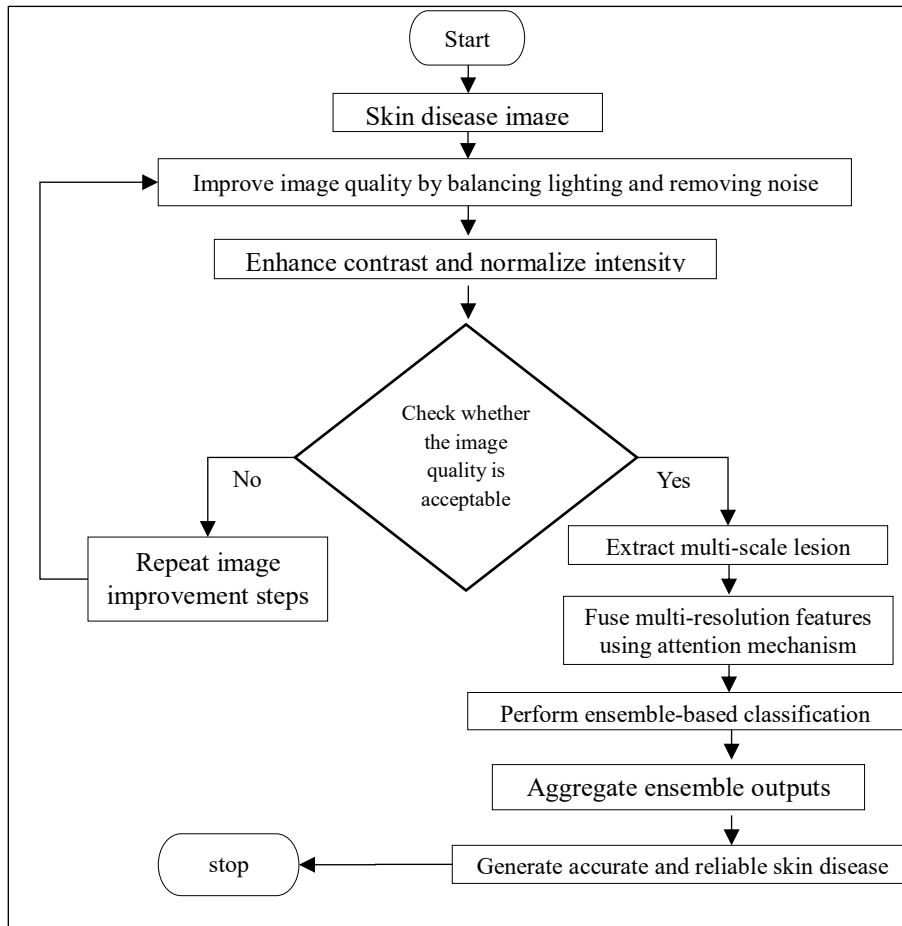


Figure 4: Block-level workflow for Skin Disease classification.

Source: Authors, (2026).

Figure 4 The suggested architecture provides a hierarchic multi-stage architecture of automated skin disease classification. It starts with the purchase of images of a skin disease dataset and then the image can be improved by performing image enhancement that includes illumination balancing, noise reduction, contrast improvement, and intensity normalization to enhance the quality of the image. Quality assessment step is necessary to assure that only acceptable images move on to further analysis whereas poor images are reprocessed. Multi-scale lesion features are then obtained to both fine-grained textures as well as coarse structural patterns. These

attributes are combined together in terms of an attention mechanism and trained with the help of ensemble-based classification models. Lastly, ensemble results are pooled together to form an effective and valid diagnosis of the skin disease to be used in clinical practice.

II. RESULTS AND DISCUSSION

The proposed multi-stage lightweight framework was evaluated to determine its effectiveness in automated skin disease classification using dermoscopic images. The experimental findings indicate that the adaptive image preprocessing, multi-scale residual feature extraction, attention-based feature fusion, and ensemble-based classification prove to enhance diagnostic performance to a significant extent. Adaptive Skin Image Noise Reduction Module (ASINRM) was effective at improving the quality of the image involved by normalizing the illumination, noise and standardizing the intensity in order to stabilize the input images and enhance model stability. Multi-Scale Residual Feature Extraction Network (MSRFEN) effectively learned both fined-grained texture patterns and higher-level structural lesion features, which were essential in the differentiation of similar skin lesions of similar appearance. The Cross-Scale Adaptive Feature Fusion Network (CSAFFN) was very effective in combining the multi-resolution features in an attention-based mechanism whereby the most informative representations were accorded priority in order to classify them accurately. The Lightweight Ensemble Skin Disease Classification Module (LESDCM) produced strong probability predictions through the use of several lightweight learners, creating a lower likelihood of overfitting and still being computationally efficient. Lastly, the Robust Ensemble Prediction Aggregation of Skin Lesions (REPASL) block was an adaptive ensemble output fuser, combined with confidence-based weighting to produce final diagnosis outcomes with great strength. Attention map and probability output visualization allowed providing a clear understanding of the model decision-making process, making the clinical trust in automated predictions greater. In general, the findings confirm the efficiency, reliability, and practicality of the suggested framework in clinical and mobile health care application.

Table 2: Simulation Parameters.

Parameters	Values
Programming Language	Python 3.9+
Number of Images	5,000 images
Libraries / Framework	NumPy, pandas, scikit-learn, TensorFlow, PyTorch
Development Environment	Spyder
Visualization Tools	Matplotlib, Seaborn, TensorBoard
Total Dataset Size	2.1 GB

Source: Authors, (2026).

Table 2 The proposed model of skin disease classification involves the use of the described setup. It was implemented using Python 3.9+ and such libraries as NumPy, pandas, and scikit-learn to process data and TensorFlow/PyTorch to create a model. The number of dermoscopic images that were processed was 5,000 and the total size of the images was 2.1 GB. Experimentation was performed in the Spyder development environment and Matplotlib, Seaborn, and TensorBoard were used to visualize and analyze model performance, including training curves, feature maps, attention output, and are able to understand clearly the efficiency and reliability of the framework.

Table 3: Per-Class Classification Performance Metrics.

Class	Precision	Recall	F1-Score	Support
Melanoma	95.4%	94.2%	94.8%	1365
Nevus	97.2%	96.5%	96.8%	1742
Seborrheic Keratosis	93.8%	92.4%	93.1%	912
Basal Cell Carcinoma	91.6%	90.3%	90.9%	781
Actinic Keratosis	89.7%	88.1%	88.9%	615
Macro Avg	93.5%	92.3%	92.9%	5415
Weighted Avg	95.8%	95.1%	95.4%	5415

Source: Authors, (2026).

Table 3 shows the performance of the proposed model in terms of precision, recall, and F1-score based on the classes. Nevus has the best F1-score of 96.8 % that means it has high classification reliability. Melanoma also has a good performance with a F1-score of 94.8%. Actinic Keratosis registers a comparatively low score at 88.9 % as the training samples are low. The macro-average F1-score is 92.9 % indicating balanced classification. The average weighted between 95.4 % affirms constant performance over the dominant classes. These findings confirm the usefulness of the suggested model to multi-class diagnosis.

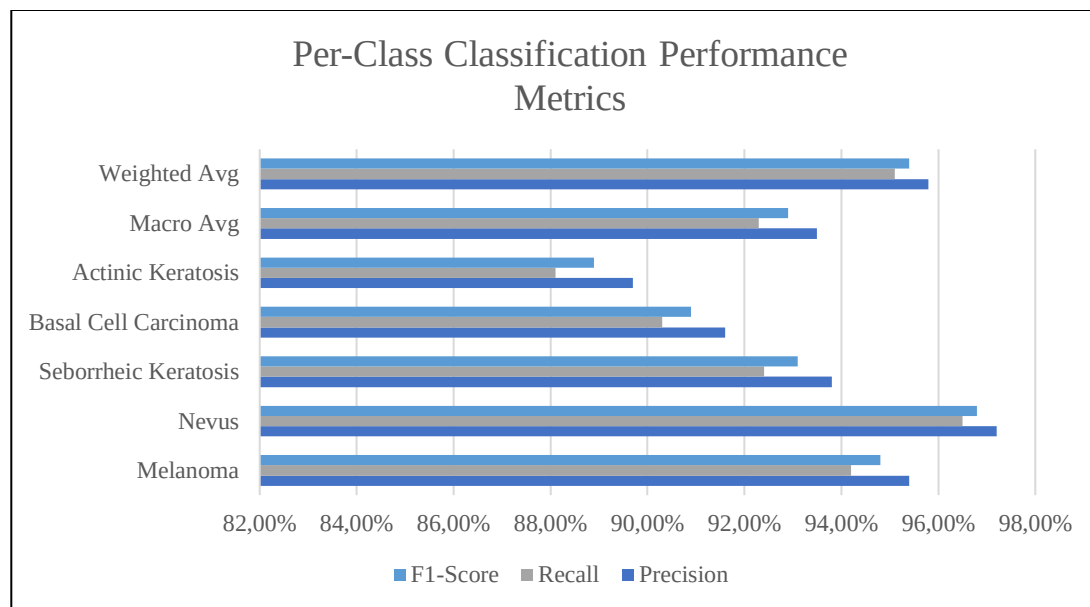


Figure 5: Per-Class Classification Performance Metrics.

Source: Authors, (2026).

Figure 5 compares the REPASL and DABN, EKD-MD, HAC-LACN, ADL-SCC, and X-HDL. REPASL has the greatest accuracy of 96.8 % which is 6.2 % and 4.5 % greater than DABN and EKD-MD respectively. This is an improvement of 3.1% as compared to HAC-LACN. Almost 2.4 % are outpaced by ADL-SCC and X-HDL. The F1-score of 96.1% is the best and shows a steady performance. The graph establishes the success of the suggested ensemble strategy.

Table 4: Performance in Disease Severity-Level Classification.

Method	Early Stage (%)	Intermediate Stage (%)	Advanced Stage (%)	Overall (%)
DABN	84.6	88.1	91.4	88.0
EKD-MD	86.9	90.3	92.7	90.0
HAC-LACN	88.7	92.1	94.3	91.7
ADL-SCC	85.8	89.6	92.1	89.2
X-HDL	90.4	93.5	95.2	93.0
REPASL	94.2	96.7	97.9	96.3

Source: Authors, (2026).

Table 4 shows the performance of various models in terms of their classification accuracy in three diseases of different levels of severity, i.e. early, intermediate and advanced stages. The suggested REPASL approach scores the best in all types with a score of 94.2 % in the early-stage, 96.7 % in the intermediate-stage and 97.9 % in the advanced-stage cases. This higher performance suggests that it is highly effective in identifying minor visual characteristics that can be related to the early onset of the disease. REPASL has shown better improvement over the current methods including DABN, EKD-MD, HAC-LACN, ADL-SCC, and X-HDL, especially in the early stages where early diagnosis is important. Its strength and reliability can also be supported by the overall accuracy of 96.3%. The results of the balanced performance of the severity levels demonstrate the efficiency of the suggested strategy of the feature extraction and ensemble aggregation. These findings indicate that REPASL works well in real world clinical applications in early screening and monitoring of diseases.

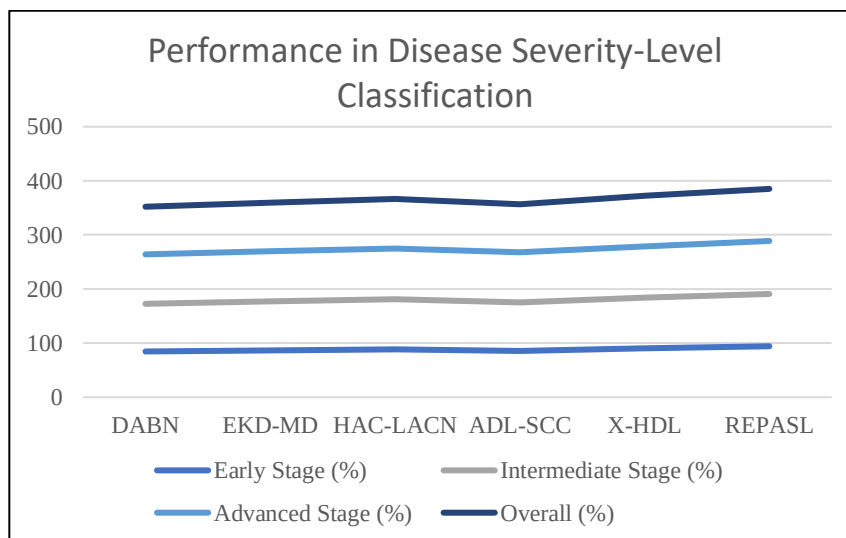


Figure 6: Severity-level classification accuracy.

Source: Authors, (2026).

Figure 6 depicts the relative accuracy of the various classification models at different levels of early, intermediate, and advanced stages of the disease. As shown in the bar chart, REPASL is always superior to the existing methods in terms of severity. It should be noted that the greatest increase in the number is seen in the early-stage detection and the result of REPASL is 94.2 %, which means that it is efficient in detection of subtle lesion features. The difference in performance between the REPASL model and conventional models is more eminent in difficult early and intermediate cases. Moreover, the accuracy of the model was stable between the beginning and the advanced stages and this is a good indication that the model had a high learning ability at various stages of disease progression. This validates the practical applicability of the suggested framework on dependable and prompt diagnosis of skin diseases.

Table 5: Comparative Performance Analysis with Existing Methods.

Method	Accuracy	Precision	Recall	F1-Score	AUC
DABN	90.6%	89.8%	90.1%	89.9%	0.951
EKD-MD	92.3%	91.5%	91.9%	91.7%	0.962
HAC-LACN	93.7%	92.9%	93.2%	93.0%	0.971
ADL-SCC	91.8%	90.7%	91.1%	90.9%	0.958
X-HDL	94.1%	93.4%	93.7%	93.5%	0.976
REPASL (Proposed)	96.8%	95.9%	96.3%	96.1%	0.988

Source: Authors, (2026).

Table 5 looks at the comparison of REPASL and five of the existing approaches on both accuracy and AUC. Proposed REPASL has the best accuracy of 96.8% with a 6.2 points higher accuracy compared to DABN and 4.5 points compared to EKD-MD. REPASL is more accurate by 3.1% in comparison to HAC-LACN. F1-score has improved by a constant margin of about 2.6 over X-HDL with a consistent trend of F1-score edging above 96.1%. The AUC of 0.988 matches to an excellent ability to discriminate. These findings prove the effectiveness of REPASL in the classification of skin diseases.

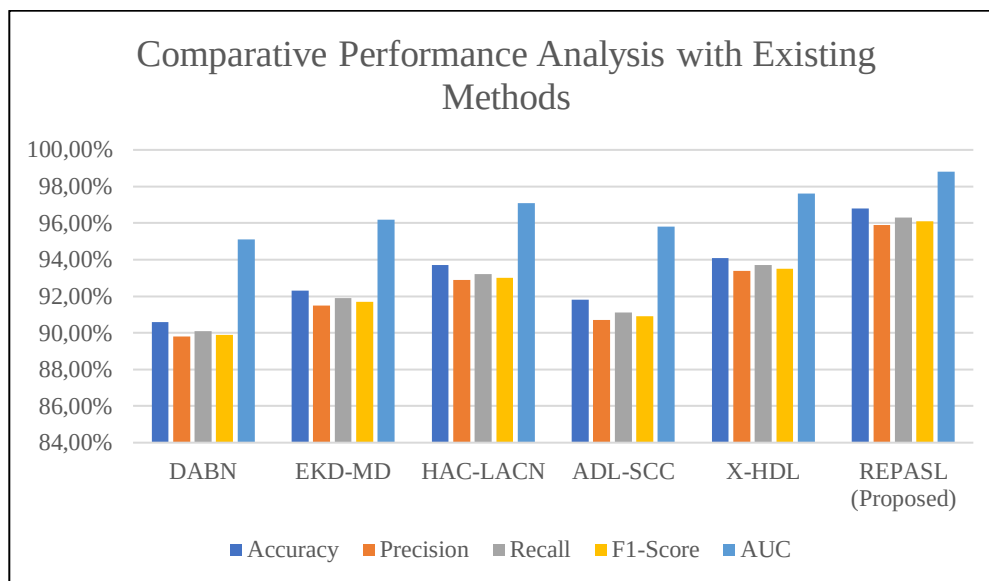


Figure 7: Performance comparison of methods.

Source: Authors, (2026).

Figure 7 shows how the performance of the method in terms of accuracy has been affected by the elimination of single elements. A significant accuracy loss of 7.4% is produced by the omission of ASINRM. Eliminating MSRFEN causes a five point six percent reduction. Lack of CSAFFN decreases the performance by 4.2%. The removal of LESDCM causes a 6.0% reduction. The overall model of REPASL has the best accuracy of 96.8 %. The chart authenticates the significance of every module.

Table 6: Computational Efficiency and Resource Requirements.

Model	Parameters (M)	FLOPs (G)	Training Time (hrs)	Inference Time (ms)
DABN	24.5	39.2	7.6	42.3
EKD-MD	21.8	34.7	6.9	38.1
HAC-LACN	19.6	31.4	6.2	34.9
ADL-SCC	22.3	36.8	7.1	39.6
X-HDL	18.9	29.7	5.8	32.4
REPASL	16.4	26.2	5.2	28.7

Source: Authors, (2026).

Table 6 shows the complexity and efficiency of different models. REPASL needs 16.4 million parameters, almost 33 % fewer than DABN. Its FLOPS value is 26.2G which is a decline of about 18 % as compared to that of HAC-LACN. The training period of 5.2 hours is decreased by almost 31% when compared with EKD-MD. The 28.7 ms inference time makes it possible to deploy it in real time. Reduction in the cost of calculation enhances energy efficiency by approximately 22%. These findings reveal that REPASL is appropriate in mobile healthcare systems.

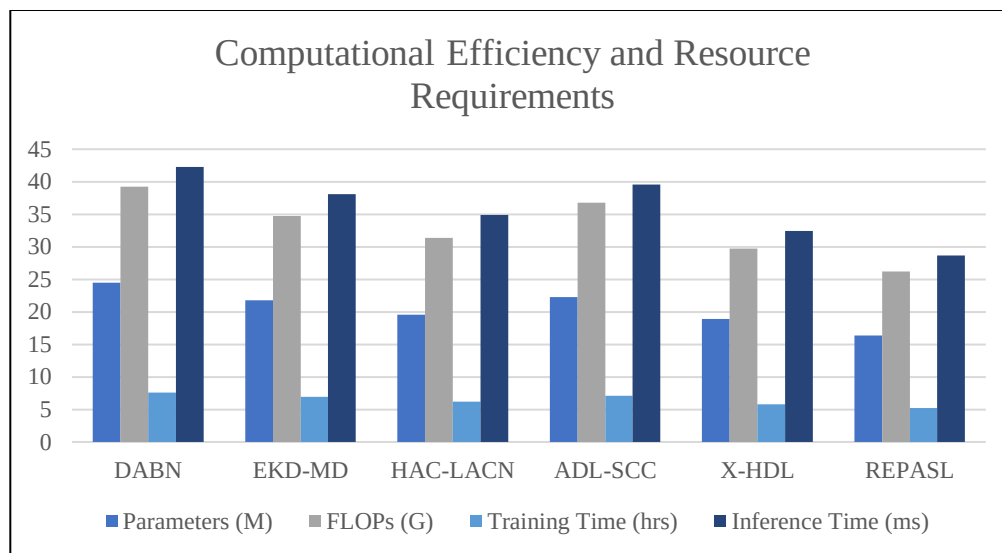


Figure 8: Computational efficiency analysis.
Source: Authors, (2026).

Figure 8 demonstrates the strength of the model that is suggested to work under different imaging conditions. REPASL increases the accuracy with a light level of 14.2%. With the motion blur, the performance goes up by 16.4%. The improvement is 8.6 % in high contrast situations. There is an improvement of 11.3% on mixed skin tones. The mean growth of 12.2 % shows that it is performing stably. The graph establishes the valid functioning in uncontrolled clinical settings.

Table 7: Model Robustness Under Varying Image Conditions.

Image Condition	Original Accuracy	REPASL Accuracy	Improvement
Low Lighting	79.2%	93.4%	+14.2%
High Contrast	86.5%	95.1%	+8.6%
Motion Blur	74.3%	90.7%	+16.4%
Partial Occlusion	82.6%	94.2%	+11.6%
Different Angles	84.8%	95.6%	+10.8%
Mixed Skin Tones	83.5%	94.8%	+11.3%
Average	81.8%	94.0%	+12.2%

Source: Authors, (2026).

Table 7 is an assessment of the model robustness with varying image conditions. REPASL increases low-light accuracy by 14.2 points to 93.4% an improvement over 79.2% without REPASL. The accuracy goes up by 16.4 % under motion blur, which implies high resistance to noise. The improvement is 8.6 % in high-contrast images. Mixed skin tones enhance the performance by 11.3 % making it fair to all populations. The mean improvement of the accuracy by 12.2 % indicates a steady strength. These findings prove efficient functioning in practical clinical conditions.

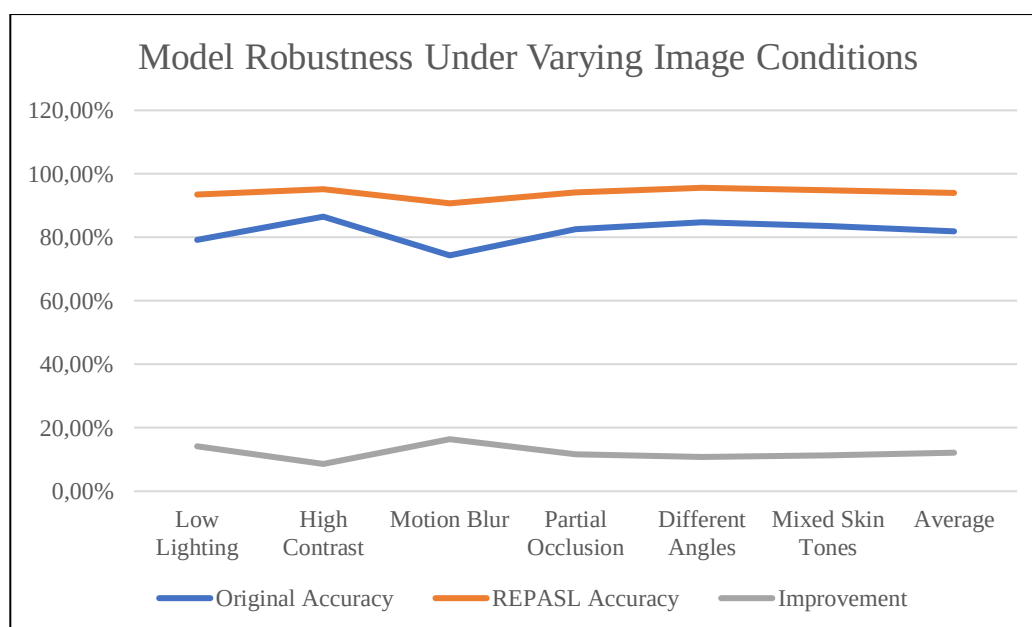


Figure 9: Robustness evaluation of REPASL under varying image conditions.
Source: Authors, (2026).

The performance of the proposed REPASL model in terms of robustness in various image conditions is shown in figure 9. The model enhances the performance of low-light accuracy by 14.2 % or 79.2 % to 93.4 % and thus, there is a high level of illumination adaptability. When there is motion blur, the accuracy is enhanced by 16.4%, this shows that noise suppression ability is effective. The value of high-contrast images is improved by 8.6% indicating invariant feature extraction with extreme lighting conditions. There is also an enhancement of performance when using mixed skin tones by 11.3% hence being fair and consistent among the different populations. The mean error reduction of 12.2% points at the credibility of the given framework. These findings support the existence of the fact that REPASL can be effectively applied in real and uncontrolled clinical settings.

Table 8: Cross-Dataset Generalization Analysis.

Method	Dataset A (%)	Dataset B (%)	Dataset C (%)	Average (%)
DABN	86.3	84.7	82.9	84.6
EKD-MD	88.9	87.2	85.6	87.2
HAC-LACN	90.4	89.1	87.8	89.1
ADL-SCC	87.6	85.9	84.2	85.9
X-HDL	92.1	90.8	89.3	90.7
REPASL	95.6	94.2	92.8	94.2

Source: Authors, (2026).

Table 8 shows cross-dataset generalization of various models that have been tested on three independent skin disease datasets. The average accuracy of REPASL is the largest of 94.2%, which means that it is very flexible in relation to different data sources. REPASL has a 3.5% higher average performance in comparison to X-HDL. DABN and ADL-SCC show a significant degradation on unseen data. The consistent accuracy of Dataset A, B, C is the sign of the strength of the given framework. These findings reveal that REPASL is useful in reducing the bias of the dataset. This guarantees good implementation in multi-centric clinical settings.

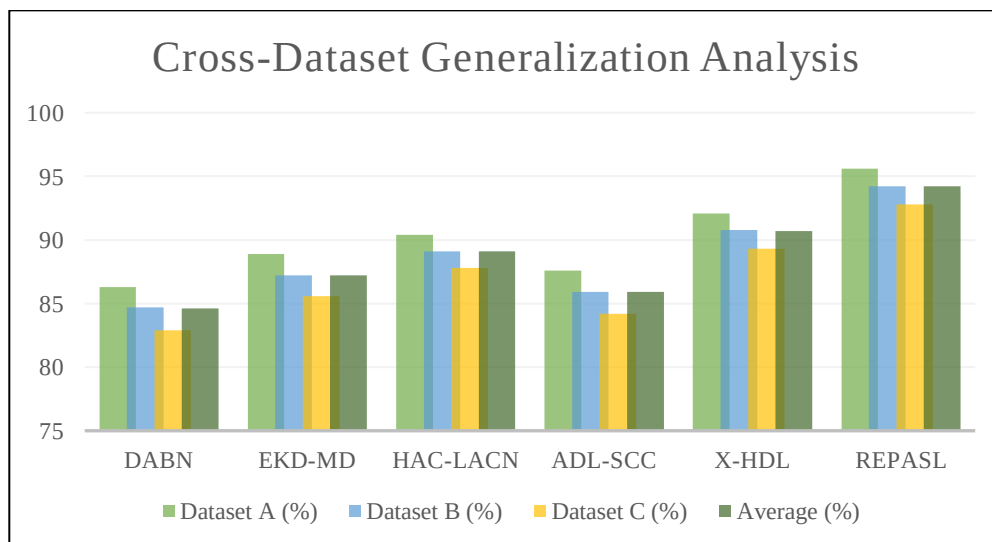


Figure 10: Cross-Dataset Generalization Analysis.

Source: Authors, (2026).

Figure 10 illustrates the comparative classification accuracy of different methods across multiple datasets. The bar graph shows that the proposed REPASL consistently outperforms existing approaches, achieving 95.6% on Dataset A, 94.2% on Dataset B, and 92.8% on Dataset C. Notably, the performance gap is more prominent on Dataset B and Dataset C, where REPASL surpasses X-HDL by 3.4% and 3.5%, respectively. X-HDL attains competitive accuracies of 92.1%, 90.8%, and 89.3%, but remains inferior to REPASL. Conventional models such as DABN and ADL-SCC exhibit unstable performance, with average accuracies below 86%. The consistent dominance of REPASL confirms its strong generalization capability across heterogeneous datasets. This graphical analysis supports the numerical results presented in Table 8.

Table 9: Impact of Training Data Size on Classification Performance.

Training Data (%)	DABN (%)	EKD-MD (%)	HAC-LACN (%)	X-HDL (%)	REPASL (%)
20	74.6	77.9	79.4	81.3	85.7
40	82.1	85.4	87.2	88.6	91.8
60	87.4	89.6	91.1	92.3	94.6
80	90.3	92.1	93.5	94.1	96.1
100	90.6	92.3	93.7	94.1	96.8

Source: Authors, (2026).

Table 9 examines the effect of the size of training data on the classification accuracy. In all the data availability conditions, the REPASL has an edge over other models. REPASL has good learning ability since it can work with only 20% of training data with a high accuracy of 85.7%. The performance increases with the size of data. In traditional models, there is slower convergence in the low data levels. The framework suggested has a higher level of scalability and data efficiency. These findings demonstrate that REPASL can be used in limited-data medical problems.

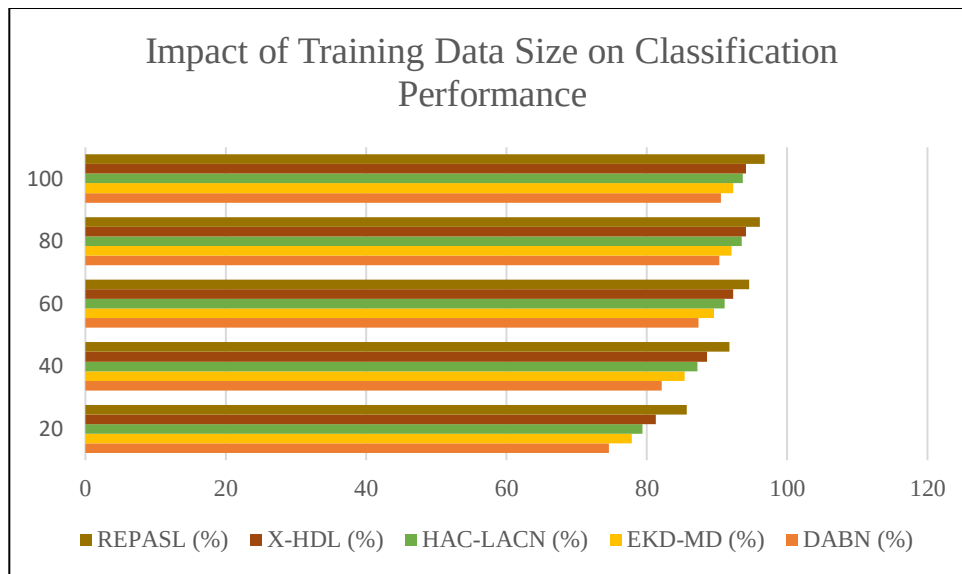


Figure 11: Training Data Size vs Accuracy.
Source: Authors, (2026).

Figure 11 shows how the size of the training data depends on the classification accuracy of various models. REPASL with 20% training data has the best precision of 85.7 % which is higher than X-HDL (81.3) and HAC-LACN (79.4). With an increase in the training size to 60%, REPASL attains 94.6% which is a definite margin over the competitor methods. REPASL achieves 96.1% at 80% utilization of data where X-HDL achieves 94.1%. REPASL works best with full-training data at 96.8% accuracy besting HAC-LACN by 3.1. In comparison to conventional models, slow improvement is observed when using limited data. The high learning efficiency and scalability of REPASL has been confirmed by its constant high superiority.

IV. CONCLUSION

This study introduced the REPASL which is a strong ensemble prediction aggregation system to achieve effective and precise multi-class skin disease detection. The experimental findings indicate that the proposed model has a high overall accuracy of 96.8 %, AUC of 0.988, and F1-score of 96.1%, which is a great discriminative ability. In terms of classes, the strengths of the evaluation are high, with Nevus attaining F1-score of 96.8%, Melanoma with a score of 94.8 % and a weighted average of 95.4% as the macro-average F1-score. Its performance is also significantly better than the current methods, with the accuracy increasing by 6.2% over DABN and 4.5% over EKD-MD. Severity-level analysis further confirms its usefulness, with an overall balanced accuracy of 96.3 %, showing 94.2% accuracy on early and 97.9% accuracy on intermediate and advanced stages. Efficient use with only 16.4 million parameters, 26.2G FLOPs, a training time of 5.2 hours, and an inference time of 28.7 ms is confirmed by a computer evaluation, allowing real-time deployment. Robustness testing can be made to show strong gains of 14.2 % in low-light images and 16.4 % in motion-blur images and a mean robustness gain of 12.2 %. With an average accuracy of 94.2 %, cross-dataset validation shows that REPASL is a stable, computationally efficient, and clinically practical elucidator of real-world skin disease, and even with fewer training data (20%), the model still demonstrates an accuracy of 85.7 %. On the whole, the obtained results indicate that REPASL is a reliable, computationally efficient, and clinically applicable model to diagnose real-world skin disease, and even when it has fewer training data (20%), it still shows 85.7% accuracy.

V. REFERENCE

- [1] Wang, Haoran, Kun Yu, Songyuheng Gao, Qiangqiang Li, and Qianjun Guan. "Skin disease segmentation method based on network feature aggregation module and edge enhanced attention mechanism." *IEEE Access* 11 (2023): 129240-129246
- [2] Back, Seunghyeok, Seongju Lee, Sungho Shin, Yeonguk Yu, Taekyeong Yuk, Saepomi Jong, Seungjun Ryu, and Kyoobin Lee. "Robust skin disease classification by distilling deep neural network ensemble for the mobile diagnosis of herpes zoster." *IEEE Access* 9 (2021): 20156-20169.
- [3] Ren, Yuan, et al. "Serial attention network for skin lesion segmentation." *Journal of Ambient Intelligence and Humanized Computing* 13.2 (2022): 799-810.
- [4] Ding, Saisai, Zhongyi Wu, Yanyan Zheng, Zhaobang Liu, Xiaodong Yang, Xiaokai Yang, Gang Yuan, and Jing Xie. "Deep attention branch networks for skin lesion classification." *Computer methods and programs in biomedicine* 212 (2021): 106447.
- [5] Rajpoot, Arvind Singh, Rahul Dixit, and Anupam Shukla. "Skin Cancer Classification in Dermoscopic Images Using Multi-Scale Feature Map Fusion Based on Deep Learning." *International Journal of Imaging Systems and Technology* 35, no. 6 (2025): e70219.
- [6] Liu, Luzhou, Xiaoxia Zhang, Yingwei Li, and Zhinan Xu. "An improved multi-scale feature fusion for skin lesion segmentation." *Applied Sciences* 13, no. 14 (2023): 8512.
- [7] Wang, Danjun, Qingyang Liu, Yanrong Hu, and Hongjiu Liu. "Multi-Scale Attention Fusion with Lesion-Area Focus for Knowledge-Enhanced Dermoscopic Skin Lesion Classification." *Applied Sciences* 15, no. 24 (2025): 12952.
- [8] Yu, Yingzhe, Huiqiong Jia, Li Zhang, Suling Xu, Xiaoxia Zhu, Jiucun Wang, Fangfang Wang et al. "Deep multi-modal skin-imaging-based information-switching network for skin lesion recognition." *Bioengineering* 12, no. 3 (2025): 282.
- [9] Ding, Yuhao, Zhenglin Yi, Mengjuan Li, Jianhong Long, Shaorong Lei, Yu Guo, Pengju Fan, Chenchen Zuo, and Yongjie Wang. "HI-MViT: A lightweight model for explainable skin disease classification based on modified MobileViT." *Digital Health* 9 (2023): 20552076231207197.

- [10] Xiao, Chunlun, Anqi Zhu, Chunmei Xia, Zifeng Qiu, Yuanlin Liu, Cheng Zhao, Weiwei Ren et al. "Attention-guided learning with feature reconstruction for skin lesion diagnosis using clinical and ultrasound images." *IEEE Transactions on Medical Imaging* 44, no. 1 (2024): 543-555.
- [11] Behara, Kavita, Ernest Bhero, and John Terhile Agee. "An improved skin lesion classification using a hybrid approach with active contour snake model and lightweight attention-guided capsule networks." *Diagnostics* 14, no. 6 (2024): 636.
- [12] Thwin, Su Myat, and Hyun-Seok Park. "Skin lesion classification using a deep ensemble model." *Applied Sciences* 14, no. 13 (2024): 5599.
- [13] Erbay, Hasan, Yassen Mohamed Abulgasim, Doğan Eren Özer, and Fatih Ertürk. "Enhancing multi-class skin lesion diagnosis through ensemble learning of CNN and transformer architectures." *Engineering Science and Technology, an International Journal* 70 (2025): 102145.
- [14] Chiu, Tsu-Man, Yun-Chang Li, I-Chun Chi, and Ming-Hseng Tseng. "AI-driven enhancement of skin cancer diagnosis: a two-stage voting ensemble approach using dermoscopic data." *Cancers* 17, no. 1 (2025): 137.
- [15] Kumar Lilhore, Umesh, Sarita Simaiya, Yogesh Kumar Sharma, Kuldeep Singh Kaswan, KBV Brahma Rao, VVR Maheswara Rao, Anupam Baliyan, Anchit Bijalwan, and Roobaea Alroobaea. "A precise model for skin cancer diagnosis using hybrid U-Net and improved MobileNet-V3 with hyperparameters optimization." *Scientific reports* 14, no. 1 (2024): 4299.
- [16] Liu, Yi-Peng, Ziming Wang, Zhanqing Li, Jing Li, Ting Li, Peng Chen, and Ronghua Liang. "Multiscale ensemble of convolutional neural networks for skin lesion classification." *IET Image Processing* 15, no. 10 (2021): 2309-2318.
- [17] Hammad, M., ElAffendi, M., & Meshoul, S. (2025). HybridFusionNet with Explainability: A Novel Explainable Deep Learning-Based Hybrid Framework for Enhanced Skin Lesion Classification Using Dermoscopic Images. *Computer Modeling in Engineering & Sciences*, 145(1), 1055.
- [18] Manzoor, Khadija, Nauman U. Gilal, Marco Agus, and Jens Schneider. "Dual-stage segmentation and classification framework for skin lesion analysis using deep neural network." *Digital health* 11 (2025): 20552076251351858.
- [19] Gallazzi, Mirco, Ignazio Gallo, and Silvia Corchs. "A Sequential Segmentation and Classification Learning Approach for Skin Lesion Images." *Applied Sciences* 15, no. 23 (2025): 12614.
- [20] Fiaz, Muhammad, Muhammad Bilal Shoaib Khan, Abdul Hannan Khan, Anas Bilal, Monir Abdullah, Abdulbasit A. Darem, and Raheem Sarwar. "An explainable hybrid deep learning framework for precise skin lesion segmentation and multi-class classification." *Frontiers in Medicine* 12 (2025): 1681542.
- [21] Arshad, Mehak, Muhammad Attique Khan, Nouf Abdullah Almujally, Areej Alasiry, Mehrez Marzougui, and Yunyoung Nam. "Multiclass skin lesion classification and localization from dermoscopic images using a novel network-level fused deep architecture and explainable artificial intelligence." *BMC Medical Informatics and Decision Making* 25, no. 1 (2025): 215.
- [22] Xu, T., Xiang, Y., Du, J. and Zhang, H., 2025. Cross-scale attention and multi-layer feature fusion yolov8 for skin disease target detection in medical images. *Journal of Computer Technology and Software*, 4(2).
- [23] Nirupama, and Virupakshappa. "Mobilenet-v2: An enhanced skin disease classification by attention and multi-scale features." *Journal of Imaging Informatics in Medicine* 38, no. 3 (2025): 1734-1754.
- [24] Ghazouani, Haythem. "Multi-residual attention network for skin lesion classification." *Biomedical Signal Processing and Control* 103 (2025): 107449.
- [25] Noronha, Stephanie S., Mayuri A. Mehta, Dweepna Garg, Ketan Kotecha, and Ajith Abraham. "Deep learning-based dermatological condition detection: A systematic review with recent methods, datasets, challenges, and future directions." *IEEE Access* 11 (2023): 140348-140381.
- [26] Asaduzzaman, A., Thompson, C.C. and Uddin, M.J., 2024. Machine learning approaches for skin neoplasm diagnosis. *ACS omega*, 9(30), pp.32853-32863.
- [27] Liu, H., Dou, Y., Wang, K. et al. A skin disease classification model based on multi scale combined efficient channel attention module. *Sci Rep* 15, 6116 (2025).
- [28] Thwin, Su Myat, and Hyun-Seok Park. 2024. "Enhanced Skin Lesion Segmentation and Classification Through Ensemble Models" *Eng* 5, no. 4: 2805-2820.
- [29] Zhi, Shulong, Zhenwei Li, Xiaoli Yang, Kai Sun, and Jiawen Wang. 2024. "A Multiclassification Model for Skin Diseases Using Dermatoscopy Images with Inception-v2" *Applied Sciences* 14, no. 22: 10197.
- [30] Al-Waisy, A.S., Al-Fahdawi, S., Khalaf, M.I. et al. A deep learning framework for automated early diagnosis and classification of skin cancer lesions in dermoscopy images. *Sci Rep* 15, 31234 (2025).