



DEMNet-CERNet: A Dual-Ensemble Framework for Automated Skin Lesion Classification Using Multiple Deep Learning Architectures

B. Subbulakshmi^{1*}, R. Divya², M. Shanmuga Priya³

¹ Department of Computer Science and Engineering, Thiagarajar College of Engineering, Madurai 625015, India

² Department of Artificial Intelligence and Data Science, Vellammal College of Engineering and Technology, Madurai 625009, India

³ Department of Computer Science and Engineering, CEG Campus, Guindy, Anna University, Chennai 600025, India

Corresponding Author Email: subbutce2015@gmail.com

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.430429>

ABSTRACT

Received: 30 May 2026

Revised: 26 July 2026

Accepted: 5 August 2026

Available online: 31 August 2026

Keywords:

skin lesion classification, deep learning, ensemble learning, transfer learning, dermoscopic images, explainable Artificial Intelligence

The global incidence of skin cancer continues to rise, with melanoma remaining the most aggressive form of disease. Deep learning (DL) has demonstrated strong performance in the automatic analysis of dermoscopic images; however, single-model architectures show limited ability to generalize to unseen dermoscopic images. Therefore, this paper presents a dual ensemble system, Deep Ensemble Model Network (DEMNet)-Convolutional Enhanced Representation Network (CERNet), for the automated identification of skin lesions. The DEMNet constructs its feature extraction pipeline using DenseNet-121, EfficientNet-B0, and MobileNetV2 to improve feature extraction, whereas CERNet employs ConvNeXt-Tiny, EfficientNetV2-B0, and ResNet50 to enhance feature representation and classification. The models use HAM10000, ISIC 2016, and ISIC 2017 datasets for training and testing. To improve model generalization and reduce the bias in predictions, various voting methods are applied to the models. The results indicate that DEMNet achieved the highest accuracy of 96% from the EfficientNet-B0 and MobileNetV2 ensemble using soft and weighted voting, while the highest accuracy achieved by CERNet is 93.9% from the ensemble of ConvNeXt-Tiny, EfficientNetV2-B0, and ResNet50 using soft voting on the HAM10000 dataset; both models exhibit high precision, recall, and F1-scores for both benign and malignant lesions. The use of heterogeneous convolutional neural network (CNN)-based ensembles with various voting methods produces improved accuracy and reliability for accurate melanoma classification. Gradient-weighted Class Activation Mapping (Grad-CAM) visualizations are employed to enhance model interpretability by highlighting the image regions that contribute most to the predictions.

1. INTRODUCTION

Melanoma is among the most lethal forms of skin cancer due to its rapid progression and ability to metastasize to other organs. Despite being less common than other types of skin cancers, late diagnosis could result in ineffective treatment and decreased chances of survival. Therefore, early and accurate diagnosis of melanoma is essential in order to facilitate the treatment process.

It is rather difficult to diagnose melanoma due to the similarity in appearance between benign and malignant skin lesions. Dermatologists usually conduct visual and dermoscopic examinations and apply the ABCDE criteria (asymmetry, border irregularity, color variability, diameter, and evolution) when evaluating suspicious lesions. However, the variability in the size, shape, color, structure, and pattern of skin lesions makes the diagnostic procedure difficult. Moreover, the presence of artifacts such as hair occlusion, noise, and uneven illumination can interfere with the identification of specific lesion features in dermoscopic images.

Large databases of dermoscopic images facilitate the application of DL techniques for skin lesion classification. HAM10000 database comprises 10,015 dermoscopic images divided into seven classes representing pigmented skin lesions and is widely used for evaluating the performance of automated classifiers [1]. Current research has demonstrated the strong capabilities of CNNs for the analysis of skin lesions. Esteva et al. [2] implemented an Inception-v3 CNN for skin lesion classification and showed performance similar to dermatologists, while Brinker et al. [3] examined the CNN classification of melanoma and benign nevi. Transfer learning using pre-trained DL models was also applied for improvement of automated melanoma detection as suggested by Kassani and Kassani [4].

Nevertheless, intra-class similarity and inter-class differences in skin lesion characteristics can complicate classification, even for individual DL models. Hence, ensemble-based methods have increasingly been adopted to take advantage of multiple models to enhance the accuracy of classification. Sharma et al. [5] introduced an ensemble that integrates both deep and manually crafted features to classify

skin lesions. Kausar et al. [6] considered the problem of multi-class skin cancer classification using modified transfer learning models, whereas Qureshi and Roos [7] proposed an ensemble framework for highly unbalanced dermoscopic data. An ensemble of multi-resolution EfficientNet models along with patient metadata was designed by Gessert et al. [8], and multi-scale image representations using ensemble learning were studied by Tsai et al. [9]. Even though DL, transfer learning, and ensemble frameworks are proven to be effective in skin lesion classification, the selection of complementary models and combining their outputs remains challenging.

This study focuses on the development of a DL-based model for classifying skin lesions through multiple pre-trained architectures and ensemble learning techniques. The method involves image pre-processing and augmentation, which enhance image quality and minimize noise and artifacts. The processed images are then fed into TL models, where pre-trained models are fine-tuned to extract discriminative features. In order to enhance the performance of classification, predictions from multiple models are combined using an ensemble learning technique. Specifically, a performance-based weighting technique has been employed to assign greater weight to models that performed well during validation.

The proposed system consists of two ensemble architectures—Deep Ensemble Model Network (DEMNet) and Convolutional Enhanced Representation Network (CERNet)—designed to capture complementary feature representations. DEMNet combines DenseNet-121, EfficientNet-B0, and MobileNetV2 to achieve dense connectivity, computational efficiency, and lightweight feature extraction. Conversely, CERNet combines ConvNeXt-Tiny, EfficientNetV2-B0, and ResNet50 to create a global contextual understanding of the entire image for both global context and hierarchical feature interpretation.

The selected models are evaluated using standard performance metrics such as accuracy, precision, recall, and F1-score to assess their performance. The present study employs Gradient-weighted Class Activation Mapping (Grad-CAM) to improve model interpretability, enhance transparency, and support clinical decision-making. This method provides a visual explanation for a model's prediction by showing which regions of interest in the input image contributed most to the model's prediction and, thereby enhancing clinical interpretability.

DL techniques, such as CNNs, TL, and ensemble DL models, have been widely employed in existing studies on skin lesion classification. Nevertheless, these approaches often exhibit limited feature representation, reduced generalizability, and lower robustness because of their reliance on individual models.

To address the issues identified by previous studies, this work proposes a novel dual-ensemble DL framework for enhancing skin lesion classification. The key contributions of this study include:

1. Introduction of a novel aggregation framework called DEMNet-CERNet, specifically designed for the automated classification of skin lesions from dermoscopic images.

2. Six contemporary deep CNN architectures are integrated within the framework, with each network contributing complementary features to improve skin lesion classification accuracy.

3. An adaptive weighting approach based on validation AUC is developed to optimally combine predictions from

multiple ensemble models.

4. Grad-CAM is employed within the proposed framework to demonstrate how the neural network interprets each classified image by visually highlighting the image regions that contribute most to lesion classification, thereby providing evidence of model interpretability.

The remainder of this paper is organized as follows. Section 2 reviews related work on CNN-based skin lesion classification, transfer learning approaches, and ensemble DL methods. Section 3 defines the proposed methodology, including image preprocessing, DL architectures, and the proposed dual-ensemble strategy. Sections 4 and 5 present the experimental setup and performance metrics. Section 6 presents the results and discussion, including comparative evaluation and Grad-CAM analysis. Finally, Section 7 concludes the paper and discusses future research directions.

2. RELATED WORKS

2.1 Convolutional neural network-based skin lesion classification

DL has become particularly effective through the development of CNNs, in which hierarchical features are learned automatically rather than being manually engineered. In many instances, CNNs have been demonstrated to produce excellent results for melanoma detection. For example, Bian et al. [10] proposed a transfer learning framework that exploits pretrained CNNs and weighted feature fusion for skin lesion classification. The proposed CNN achieved more accurate melanoma detection than traditional machine learning techniques based on handcrafted features. Juan et al. [11] proposed a deep CNN incorporating a fusion strategy and lifelong learning to classify five categories of skin cancer from clinical images. Their work achieved dermatologist-level performance, demonstrating improved accuracy, sensitivity, and robustness for automated skin cancer diagnosis using a relatively small clinical image dataset. Numerous other works have established the viability of CNNs for the recognition of melanoma. Nawaz et al. [12], for example, presented a skin cancer detection framework that integrates fuzzy k-means clustering for lesion separation with DL-based feature extraction and classification. The integration of accurate lesion segmentation and CNN-based classification improved melanoma detection performance. Toprak and Aruk [13] proposed a hybrid CNN that combines complementary feature extraction capabilities from multiple CNN architectures for multi-class skin cancer classification.

2.2 Transfer Learning approaches

TL has emerged as a highly effective approach for enhancing the performance of skin lesion classifiers, particularly in situations where large labeled medical datasets are unavailable or limited. TL enables the reuse of learned feature representations, thereby bridging the gap between DL models trained on large-scale datasets (e.g., ImageNet) and the limited availability of labeled dermoscopic image datasets. Using this technique, Venugopal et al. [14] proposed a TL-based system that utilizes the modified EfficientNet model with feature extraction optimization for automatic skin cancer detection using dermoscopic images. The enhanced EfficientNet architecture is effective for early and reliable skin

lesion diagnosis, reducing computational complexity while improving classification accuracy. Likewise, the study by Shakya et al. [15] provided an overview of the recent developments in DL and TL for skin cancer detection, highlighting the effectiveness of CNNs, Vision Transformers, and ensemble models in improving diagnostic accuracy.

Meanwhile, in a similar study, Manole et al. [16] developed an EfficientNet-based system for automatic skin lesion detection using dermoscopic images. The model leveraged TL to enhance diagnostic performance. However, TL models remain susceptible to limited training data diversity and overfitting, particularly when using small-scale medical datasets as training sets.

Vieira et al. [17] conducted a comprehensive review of DL approaches for skin lesion detection, including CNN-based models, Vision Transformers, hybrid frameworks, and ensemble approaches, using benchmark datasets such as ISIC and HAM10000, with hybrid and transformer-based models demonstrating superior diagnostic accuracy and interpretability.

Cherukuri et al. [18] proposed EnsembleSkinNet, a TL-based ensemble framework that combines Modified VGG16, ResNet50, InceptionV3, and DenseNet201 with Bayesian hyperparameter optimization and Grad-CAM for explainable skin cancer diagnosis. The framework is limited by its evaluation on a single dataset, high computational complexity, and the lack of external clinical validation.

2.3 Ensemble deep learning approaches

Ensemble learning strategies have become increasingly popular for developing more robust and reliable DL models for image classification. These methods aggregate predictions from multiple models to capture complementary feature representations, hence reducing the variance amongst the predictive models. Tsai et al. [9] proposed a multi-model ensemble that combines predictions from multiple CNNs in order to make use of complementary feature representations for skin lesion classification. In another work, Qureshi and Roos [7] used TL to develop an ensemble of multiple deep neural networks in order to classify skin cancer.

V et al. [19] proposed an ensemble of pretrained CNN models, demonstrating the effectiveness of weighted model combination for skin lesion classification.

A stacking-based framework that integrates deep features from multiple pretrained architectures was suggested by Chand et al. [20]. Although these studies demonstrate the benefits of combining complementary deep representations, they primarily focus on conventional weighted ensembles, feature-level fusion, or stacking-based meta-learning.

Sethanan et al. [21] proposed the Double AMIS-Ensemble framework to improve skin cancer classification through an optimized ensemble strategy. Chatterjee et al. [22] introduced the IncepX-Ensemble model, which combines TL with ensemble learning for early multiclass skin lesion detection. Chanda et al. [23] introduced DCENSnet, a deep convolutional ensemble network that combines three customized CNN models with different dropout configurations to improve multiclass skin lesion classification. The suggested framework attained improved results on the HAM10000 dataset but requires validation on more diverse clinical datasets.

Although ensemble learning shows considerable potential, most existing frameworks rely on a limited number of CNN architectures and are typically trained on a single dataset, thereby restricting their ability to generalize beyond the original training data.

2.4 Explainable Artificial Intelligence for skin lesion analysis

Explainable Artificial Intelligence (XAI) plays a crucial role in healthcare because it enables clinicians to understand the reasoning behind a model's diagnostic predictions. Consequently, numerous studies have focused on skin lesion classification using XAI techniques. One of the most widely used XAI techniques in skin lesion classification is Grad-CAM, which generates heat maps highlighting the image regions that are most relevant to a model's prediction. These heat maps help clinicians understand which image regions influence the model's predictions and visualize the lesion characteristics considered important for classification.

In addition to Grad-CAM, some recent studies have examined other ways to provide explanations for the classifications made by the DL models when applied to dermoscopic images, including attention-based mechanisms and other explainability frameworks.

Table 1. Comparison of existing skin lesion classification methods

Study	Methodology	Key Contribution	Research Gap/Limitation
Esteva et al. [2]	Inception-v3 CNN trained on large clinical image repository	Demonstrated dermatologist-level skin lesion classification using deep CNNs.	Requires a very large annotated clinical dataset and limited applicability to computationally efficient ensemble learning.
Kausar et al. [6]	Ensemble of multiple fine-tuned CNNs	Improved classification through multi-model feature learning.	Large ensemble increases computational complexity and does not consider performance-aware model contribution.
Qureshi and Roos [7]	Transfer Learning ensemble of six CNNs	Combined multiple pre-trained networks for improved prediction.	Large number of models increases inference time and computational complexity.
Tsai et al. [9]	Multi-model DenseNet ensemble	Exploited ensemble learning to improve robustness.	Depends on synthetic image generation and incurs high computational cost.
Juan et al. [11]	Fusion-based DCNN	Combined information from multiple clinical datasets.	Limited validation on benchmark public datasets and moderate generalization.
Nawaz et al. [12]	Region-based CNN	Integrated lesion localization and classification.	Focuses primarily on lesion detection and segmentation rather than optimizing classification performance.
Manole et al. [16]	EfficientNetB3-based custom CNN	Demonstrated strong CNN performance on dermoscopic images.	Sensitive to class imbalance and limited minority-class representation.

Note: CNN = Convolutional neural network; DCNN = Deep Convolutional Neural Network.

Dakhli and Barhoumi [24] developed an ensemble framework with explainability based on Dempster–Shafer evidence fusion, improving both classification performance and model interpretability.

As a result of using XAI techniques, improvement in the interpretability of DL models is achieved, making AI systems more likely to be adopted in a clinical practice setting.

2.5 Limitations

As can be seen from the current literature, methods such as DL, transfer learning, and ensembles have been proven successful in classification of skin lesions. Nevertheless, several limitations remain, including high computational complexity of large ensembles, susceptibility to imbalanced classes, reliance on artificial data or large datasets, and absence of performance-oriented fusion. Table 1 presents the methodology, contribution, research gap, and motivation behind the proposed framework.

As shown in Table 1, past works have focused on individual CNN architectures, transfer learning, feature fusion, and multi-model ensemble, but still face challenges in trading off classification performance against efficiency. Large ensembles might incur additional computation cost, and existing fusion methods may not adequately account for the individual performance of each model. These observations inspired the proposed dual-ensemble approach that combines complementary CNN models along with image pre-processing

and class-weighted learning, together with validation-based probability fusion, to perform robust and efficient melanoma classification.

3. PROPOSED METHODOLOGY

3.1 Overview of proposed architecture

The present study introduces a novel Dual Ensemble Melanoma Classification Framework, DEMNet-CERNet, for accurate and robust skin lesion classification using dermoscopic images. The suggested framework assimilates multiple deep CNNs organized into two complementary model groups and combines their predictions through an ensemble strategy. The main objective is to enhance classifier performance by utilizing complementary feature representations generated by different DL models. The framework also enables visualization of the classification decisions made by each model using explainable AI techniques.

Figure 1 illustrates the architecture of the proposed DEMNet-CERNet framework. It consists of five major stages: input data acquisition, data preprocessing, dual-model feature extraction, ensemble classification, and explainable prediction visualization. These stages collectively enable the system to perform reliable melanoma detection across heterogeneous dermoscopic datasets.

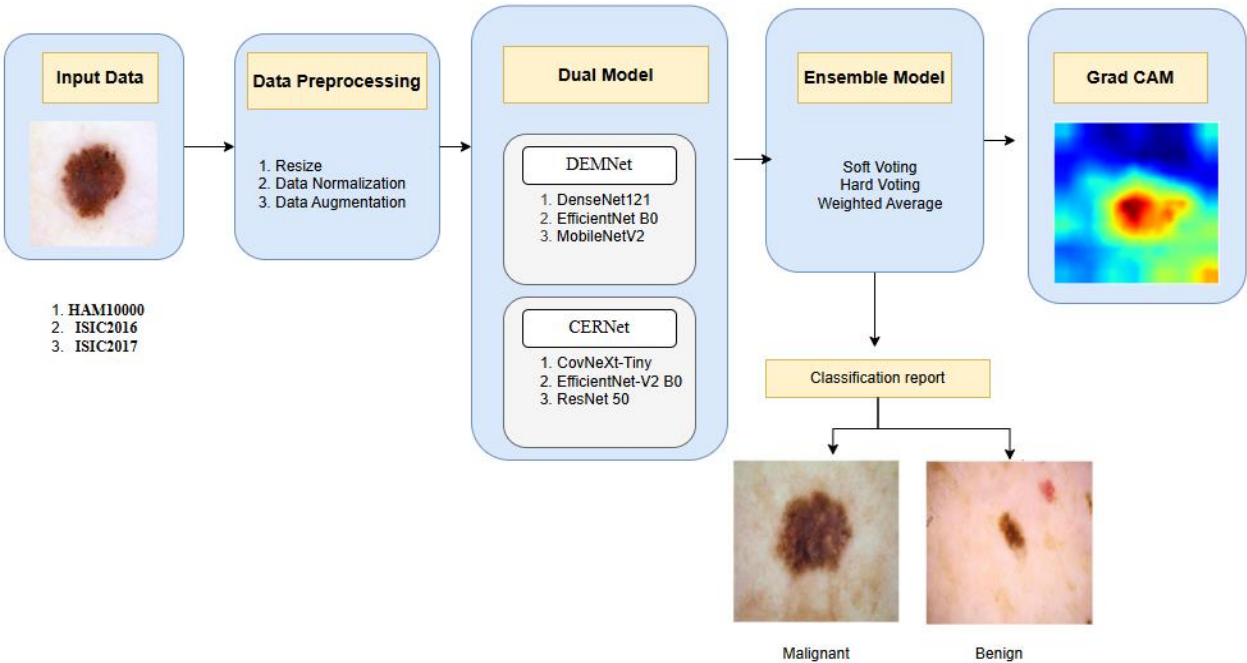


Figure 1. Proposed Deep Ensemble Model Network (DEMNet)-Convolutional Enhanced Representation Network (CERNet) framework

3.2 Dataset description

The development of DL models for skin lesion classification has been significantly facilitated by the availability of publicly accessible dermoscopic image datasets. Among these datasets, HAM10000 is one of the largest and most widely used resources for skin lesion classification. It contains more than 10,000 dermoscopic images from various clinical facilities and includes images from each pigmented skin lesion class. The large size of the

dataset makes it valuable for training DL models. Similarly, the ISIC 2016 and ISIC 2017 challenge datasets have played a significant role in advancing automated melanoma detection research. The ISIC challenge datasets come from the ISIC international challenge; both datasets contain dermoscopic images that have been accurately annotated for melanoma classification and lesion segmentation. In addition to providing annotated images, the ISIC challenge establishes standardized benchmarks for evaluating algorithm performance, enabling fair comparisons among different

algorithms and supporting the development of robust automated skin lesion classification systems. It is essential that models be trained on multiple different image datasets prior to evaluation. This is due to variations in image acquisition conditions, illumination, and lesion-class distributions that will be encountered in a variety of clinical environments within dermatology.

The proposed dual-ensemble-based methodology was evaluated on three benchmark publicly available dermoscopic image datasets comprising diverse skin lesion types, expert-annotated labels, and widely recognized standards for the development and evaluation of automated DL-based skin lesion classifiers. These datasets are HAM10000, ISIC 2017, and ISIC 2016. Table 2 summarizes the datasets used for melanoma classification, along with sources, number of images, and their classification tasks.

3.2.1 HAM10000 dataset

HAM10000 is a widely used benchmark dataset for skin lesion classification. The HAM10000 dataset contains 10,015 dermoscopic images collected from various clinical sources, including the Department of Dermatology of the Medical University of Vienna, the Skin Cancer Practice in the state of Australia, etc. The database comprises 7 different types of pigmented skin lesions.

For this study, the HAM10000 dataset was converted from a multiclass dataset into a binary classification dataset. Images labeled as melanoma (mel) were assigned to the positive class, whereas all other lesions were assigned to the non-melanoma class. This conversion enables the development of DL models specifically designed for melanoma detection. Because of this, the HAM10000 dataset's size and diversity will continue to serve as a benchmark for the evaluation of automated skin lesion classification systems.

3.2.2 ISIC 2016 dataset

The ISIC 2016 dataset, part of the ISIC Skin Lesion Analysis Challenge, was created to provide a collection of dermoscopy images to facilitate the development of automated melanoma detection algorithms using both clinical dermoscopy images and clinical expert annotations. The primary focus of this dataset is the binary classification of melanoma and non-melanoma lesions based on images collected from multiple clinics around the world. The ISIC 2016 dataset has been an important resource in DL research for the evaluation of melanoma detection models.

3.2.3 ISIC 2017 dataset

The ISIC 2017 dataset is an enhanced version of the ISIC Challenge Dataset and includes lesion segmentation, dermoscopic feature extraction, and disease diagnosis. The dataset comprises three lesion categories: melanoma, seborrheic keratosis, and benign nevi. The ISIC 2017 dataset contains high-quality dermoscopic images with corresponding ground-truth annotations. This enables the development and evaluation of advanced DL methods for both image segmentation and classification tasks. Numerous studies have used this dataset to improve automated melanoma identification systems because the ISIC 2017 dataset contains both annotated lesion masks and diagnostic classifications.

The dataset was converted to a binary classification problem in line with the research objective of detecting melanoma lesions. Melanomas were classified under the positive class, while seborrheic keratosis and benign nevi were categorized under the negative class (non-melanomas). This was chosen in order to allow the model to learn how to classify melanomas against non-melanomas that have a visually similar appearance.

Table 2. Description of datasets

References	Dataset	Number of Images	Classes	Task
[1]	HAM10000	10015	7	Binary classification
[25]	ISIC 2016	1,279	2	Binary classification
[25]	ISIC 2017	2,000	3	Binary classification

3.3 Data pre-processing

Image preprocessing is an essential step because raw medical images may contain variations in scale and orientation, as well as background interference. Pre-processing dermoscopic images allows for better feature extraction and assists in building improved DL models. Many studies have demonstrated that preprocessing can improve the performance of DL models for melanoma detection. Therefore, before training the DEMNet-CERNet dual-ensemble framework, various preprocessing techniques were applied. One of the limitations of the current work is that the pre-processing does not include dedicated hair removal or image inpainting techniques, and these techniques will be considered for future work to improve the classification performance for images with significant visual artifacts.

3.3.1 Image resizing

The spatial resolution of images varies across the benchmark datasets. To provide a uniform input size for the DL architectures, all images are resized to 224×224 pixels, which is a commonly used input size for architectures such as DenseNet, EfficientNet, and MobileNet.

3.3.2 Data augmentation

The scarcity of annotated dermoscopy images presents a major problem while training the DL models. In order to augment the diversity of the dataset and to avoid overfitting, data augmentation methods have been used during the training of the algorithm. Let the original image be denoted as I . The augmented image I' is generated using transformation function $T : I' = T(I)$. A variety of image augmentation operations were incorporated during training to enhance data diversity.

Random Flip. Horizontal and vertical flipping simulate variations in lesion orientation and improve model invariance.

Random Rotation. Dermoscopic lesions may appear at arbitrary angles. Therefore, random rotation was applied to improve rotational invariance.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (1)$$

where, θ represents the rotation angle. Rotation-based augmentation has been shown to significantly improve melanoma classification accuracy in CNN models.

Random Zoom. Random zoom simulates variations in lesion scale and size.

$$I'(x, y) = I(sx, sy) \quad (2)$$

where, s denotes the scaling factor that controls the degree of zoom with

$$s \in [0.9, 1.1]$$

Values below 1 represent a slight zoom-out, whereas values above 1 represent a zoom-in. This transformation improves the ability of the DL models to recognize lesions across variations in scale.

3.3.3 Image normalization

Before being fed into the neural network models, the pixel values of the images were normalized to facilitate faster convergence and more stable training. For an image $I(x, y)$, pixel normalisation is expressed mathematically as:

$$I_n(x, y) = \frac{I(x, y)}{255} \quad (3)$$

This operation takes pixel values from the range $[0, 255]$ and converts them into their equivalent normalised pixel value $[0, 1]$. Pixel normalization is commonly employed in DL pipelines to standardize input data, thereby improving training stability, facilitating faster convergence, and enhancing optimization performance.

3.3.4 Dataset splitting

To develop the models and evaluate their performance objectively, the dermoscopic images were randomly divided into three non-overlapping sets: training, validation, and test sets. The distribution was chosen as 80:10:10 for the training, validation, and test sets, respectively. In particular, this splitting strategy helps prevent data leakage, facilitates model tuning during training, and enables evaluation of the generalization performance of the developed models.

3.3.5 Class balancing

Typically, there is a class imbalance in skin lesions, with the benign class having a very large number of examples compared to the malignant class. Such class imbalance can bias the model toward predicting the majority class. Class weighting was used to eliminate this effect during training. Let N = total number of samples, N_i = number of samples in class i , C is the number of classes.

The class weight is computed as

$$W_i = \frac{N}{C \times N_i} \quad (4)$$

Applying class weights increases the penalty for misclassifying minority-class samples and is intended to improve the classification of melanoma lesions.

3.4 Dual model feature extraction

The major component of the proposed framework is a dual-ensemble architecture comprising two complementary modules, DEMNet and CERNet, each consisting of multiple CNN architectures that extract diverse and discriminative feature representations from dermoscopic images. The two models are referred to as DEMNet and CERNet. Each of these

two models contains multiple different CNNs that capture different levels of melanoma characteristics to produce features for skin lesions based on spatial hierarchies and semantics. There have been substantial studies demonstrating that DL-based feature extraction can automatically learn discriminative representations from dermoscopic images. However, there are limitations in using a single training architecture, particularly the poor ability to generalize across heterogeneous datasets. To avoid these limitations, the proposed approach utilizes multiple CNN architectures to learn complementary features of a lesion, thus being able to create more robust classification.

3.4.1 Deep Ensemble Model Network module

The DEMNet focuses on extracting rich hierarchical image features using well-established CNN architectures. TL is applied to leverage pretrained models trained on large-scale datasets such as ImageNet, which improves convergence and feature representation for medical imaging tasks.

DEMNet integrates the following CNN architectures: DenseNet-121, EfficientNet-B0, and MobileNetV2. Each architecture contributes unique structural characteristics that improve feature diversity.

DenseNet-121 has very dense connectivity between layers, which allows for the reuse of features above and below through efficient gradient propagation. One of the reasons for DenseNet's effectiveness in medical image classification is its ability to capture very fine-grained spatial information.

EfficientNet-B0 employs a compound scaling approach that automatically adjusts the network's width, depth, and image resolution, leading to improved performance using fewer parameters. EfficientNet models have demonstrated strong performance in skin lesion classification.

MobileNetV2 is a lightweight architecture designed using depth-wise separable convolutions and inverted residual blocks. Despite its low computational complexity, MobileNetV2 provides effective feature learning capabilities suitable for medical image classification.

This framework makes use of TL through fine-tuning of pre-trained CNNs for the classification of dermoscopic images.

Before model training, image preprocessing and lesion enhancement are performed to improve image quality. Subsequently, image augmentation techniques, including rotation, flipping, and scaling, are applied to increase data diversity and reduce overfitting. The processed images are then fed into the CNN architectures to extract deep feature representations for melanoma classification.

3.4.2 Convolutional Enhanced Representation Network module

CERNet is designed to further improve feature diversity by incorporating modern convolutional architectures capable of capturing complex visual patterns in dermoscopic images. CERNet integrates the following DL models: ConvNeXt-Tiny, EfficientNetV2-B0, and ResNet50.

ConvNeXt-Tiny employs modern convolutional design patterns influenced by transformer architectures to enhance how features are represented and to allow the model to scale.

EfficientNetV2-B0 improves training efficiency through progressive learning and optimized scaling, thereby enabling faster training and strong performance in medical image classification.

ResNet50 has been designed with residual connectivity

between layers to overcome the vanishing gradient issue that occurs with DNNs. In addition, residual learning facilitates the learning of complex visual features and improves training stability, making ResNet50 suitable for skin lesion classification.

The outputs provided by the DEMNet and CERNet modules contain prediction probabilities for skin lesion classes; these probabilities are ultimately combined using ensemble learning techniques.

The models for DEMNet and CERNet are selected to provide diverse and complementary CNN architectures. DenseNet-121 and ResNet50 capture hierarchical features through dense and residual connections; EfficientNet-B0 and EfficientNetV2-B0 provide efficient feature extraction through compound scaling and optimized training; MobileNetV2 offers computational efficiency through depthwise separable convolutions; and ConvNeXt-Tiny introduces modern CNN design principles inspired by transformer architectures. This diversity enables the proposed framework to utilize complementary feature representations and enhance classification robustness. Since this work focuses on CNN-based frameworks, transformer-based models are not considered and are left for future investigation.

3.5 Ensemble classification strategy

The dual-model system uses ensemble learning techniques to combine the predictions made by each of the models in the dual system. Ensemble learning can improve classification accuracy by aggregating the outputs of multiple models, thereby reducing prediction variance and enhancing the robustness of the resulting model.

The following ensemble techniques are implemented: Soft Voting, Hard Voting, and Weighted Averaging.

3.5.1 Soft voting

Soft voting combines the predicted class probabilities from all models and assigns the class with the highest average probability as the final prediction. If $P_i(x)$ is the probability that model i will predict the label for input image x , the result of soft voting for the ensemble of N models is:

$$P_{soft}(x) = \frac{1}{N} \sum_{i=1}^N P_i(x) \quad (5)$$

The final predicted class is obtained as:

$$y_{pred} = \arg \max(P_{soft}(x)) \quad (6)$$

Soft voting allows the ensemble to leverage the confidence levels of individual models, improving prediction stability.

3.5.2 Hard Voting

Hard Voting selects the final predicted class label by choosing the majority predicted class label from each of the models in the ensemble. If $y_i(x)$ is the predicted class label returned by model i for input image x , the final predicted class label by the ensemble of N models is:

$$y_{pred} = \text{mode}(y_1(x), y_2(x), \dots, y_N(x))$$

This strategy is effective when multiple models produce consistent predictions, reducing the impact of occasional misclassifications.

3.5.3 Weighted average ensemble

The weighted averaging strategy aggregates the outputs of the constituent models by assigning performance-based weights derived from the validation set. Let w_i represent the weight assigned to model i , where

$$\sum_{i=1}^N w_i = 1 \quad (7)$$

The weighted ensemble prediction is calculated as:

$$P_{weighted}(x) = \sum_{i=1}^N w_i P_i(x) \quad (8)$$

The final prediction is obtained as:

$$y_{pred} = \arg \max(P_{weighted}(x)) \quad (9)$$

Unlike traditional ensemble models, in which all models contribute equally to the ensemble decision, the proposed framework uses a PAF strategy to combine the predictions of multiple models. Each model was assigned a weight based on its validation AUC, so that models with better validation performance contributed more to the final prediction. The weight for the i^{th} model was computed as:

$$w_i = \frac{AUC_i}{\sum_{i=1}^N AUC_i} \quad (10)$$

where, AUC_i is the validation AUC of the i^{th} model, N is the number of models in the ensemble, and the sum of weights should be equal to 1. The final prediction was obtained by a weighted average of the class probability outputs.

This PAF mechanism improves classification robustness by prioritizing models demonstrating stronger validation performance.

3.6 Explainable Artificial Intelligence

In order to improve the model interpretability, the proposed framework applies a technique known as Grad-CAM. This technique generates visual heat maps that highlight regions of the image that contribute most strongly to the model's prediction.

Given the feature maps A^k from the final convolutional layer and the gradient of the target class score y^c , Grad-CAM computes importance weights as:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (11)$$

The final Grad-CAM heatmap is generated as:

$$L_{Grad-CAM}^c = \text{ReLU}(\sum_k \alpha_k^c A^k) \quad (12)$$

Grad CAM provides a visual representation of the lesion-relevant areas in the dermatoscopic image, which allows clinicians to review how the model reached its final label. Additionally, misclassified examples have been incorporated into the qualitative analysis to provide a more balanced interpretation of the model. These examples demonstrate that incorrect predictions may occur when lesions exhibit visual similarities across classes or have ambiguous boundaries and low contrast, as well as in the presence of artifacts such as hair and variations in image acquisition.

4. EXPERIMENTAL SETUP

The experiments were conducted in a DL environment using Python and the TensorFlow and Keras libraries. Models were trained and tested on Google Colab Pro with an NVIDIA Tesla T4 GPU for accelerated computations in DL applications. Prior to training, all images were resized to 224×224 pixels. In order to avoid overfitting and provide enough iterations to learn features efficiently, the models were built by applying a batch size of 32 over 50 epochs. For gradient optimization, the Adam optimizer with a learning rate of 1×10^{-3} in the first training stage and 1×10^{-4} in the fine-tuning stage was used to provide stable training and increase the performance of the classification task. The model parameters were optimized using the categorical cross-entropy loss function. During training, data augmentation was performed by random flipping, rotation, zooming, and contrast adjustment. Moreover, class-weighted training was employed to mitigate the effects of class imbalance, and the random seed was set to 42.

Code Availability: The source code supporting the findings of this study is available from the corresponding author upon reasonable request.

5. PERFORMANCE METRICS

Several performance metrics, including accuracy, precision, recall, F1-score, and specificity, were used to evaluate the developed framework. Accuracy refers to the overall correctness of predictions, while Precision and recall reflect the model's ability to correctly identify melanoma lesions. F1-score represents the balance between precision and recall,

while specificity is concerned with the capacity of the model to recognize benign lesions. Therefore, the models' discriminative abilities were evaluated at different decision thresholds through the use of an ROC curve and an AUC score. These metrics provide a comprehensive evaluation of the performance of DEMNet, CERNet, and the proposed dual-ensemble framework for skin lesion classification.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

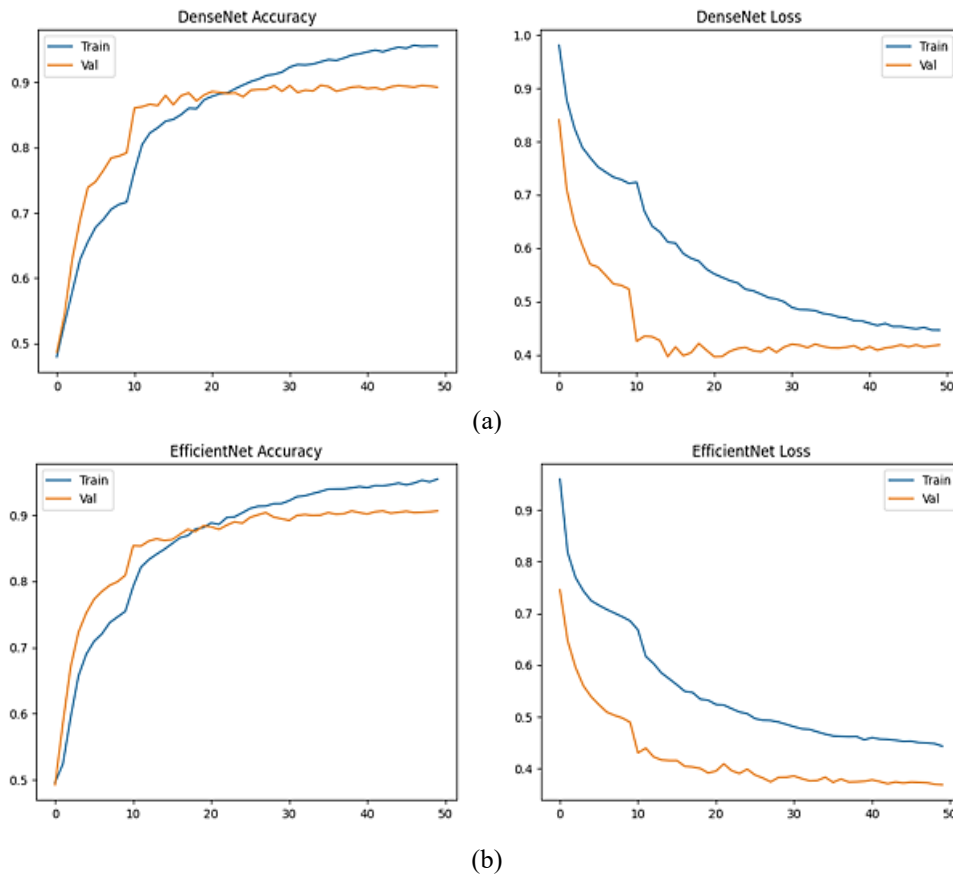
$$Precision = \frac{TP}{TP + FP} \quad (14)$$

$$Recall = \frac{TP}{TP + FN} \quad (15)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (16)$$

6. RESULTS AND DISCUSSION

This section presents the experimental results obtained using the proposed DEMNet-CERNet framework for skin lesion classification. The evaluation is done based on the use of various metrics that include accuracy, precision, recall, and F1-score. The experiments were conducted on the dermoscopic images of the HAM10000, ISIC 2017, and ISIC 2016 datasets. The proposed framework integrates multiple deep CNNs, including DenseNet-121, EfficientNet-B0, MobileNetV2, ConvNeXt-Tiny, EfficientNetV2-B0, and ResNet50, through ensemble learning techniques.



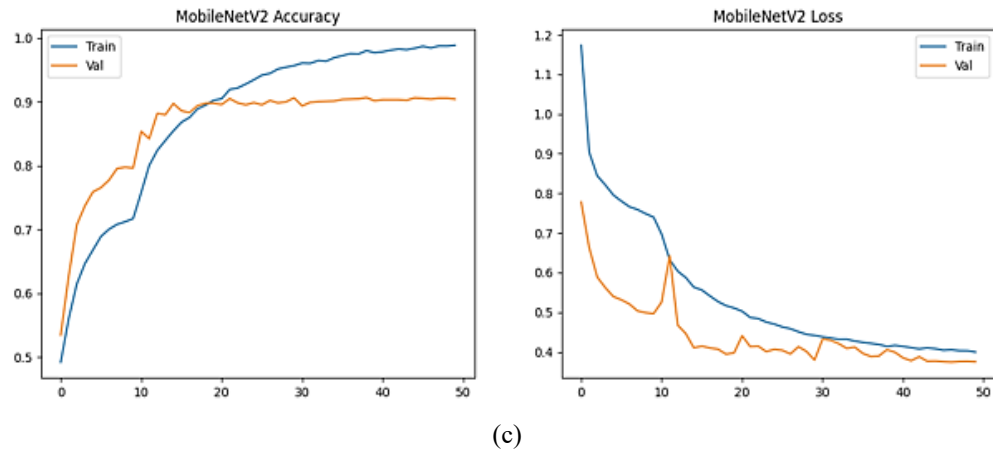


Figure 2. Accuracy and loss trends of the Deep Ensemble Model Network (DEMNet) module during training and validation on the HAM10000 dataset, (a) DenseNet-121, (b) EfficientNet-B0, and (c) MobileNetV2

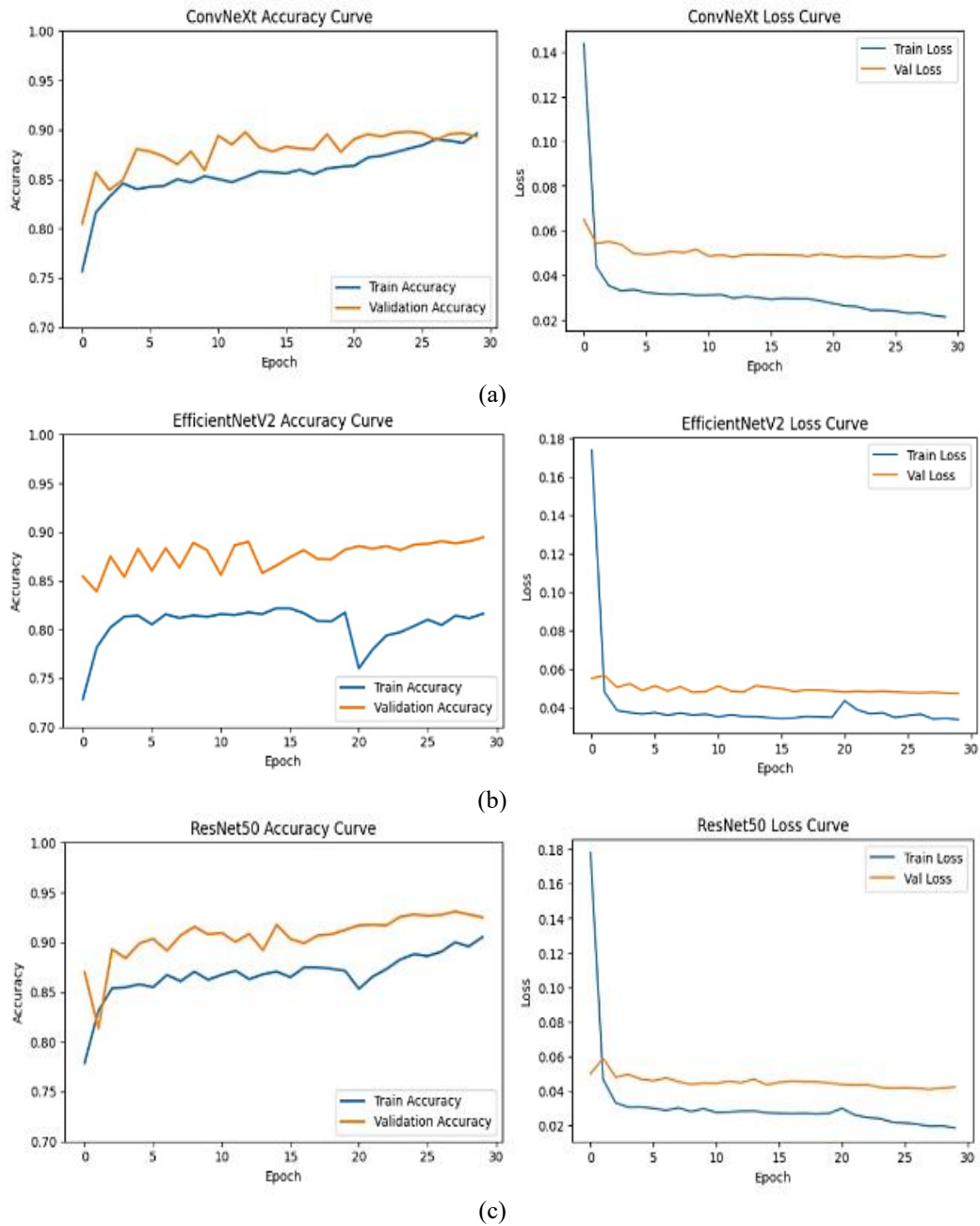


Figure 3. Accuracy and loss trends of the Convolutional Enhanced Representation Network (CERNet) module during training and validation on the HAM10000 dataset, (a) ConvNeXt-Tiny, (b) EfficientNetV2-B0, and (c) ResNet50

The training and validation performances of three different CNNs utilized in the DEMNet module, DenseNet-121, EfficientNet-B0, and MobileNetV2, on the HAM10000 dataset are illustrated in Figure 2. Curves depicting accuracy and loss over 50 epochs show that as the training and validation accuracies increase over time, the losses decrease. This trend indicates that the models progressively learned discriminative representations for skin lesion classification and achieved convergence during training. Based on the models tested, MobileNetV2 achieved the highest level of training accuracy while EfficientNet-B0 achieved the best validation score; therefore, they would be expected to provide excellent capability for extracting high-quality deep features from clinical images used for classifying melanomas.

Figure 3 illustrates the accuracy and loss values of the CERNet module during training and validation on HAM10000 dataset, which integrates three DL backbones: ConvNeXt-

Tiny, EfficientNetV2-B0, and ResNet50. The plots show a constant improvement in training performance across epochs, while the validation accuracy stabilizes around a high range, indicating good generalization capability. The loss curves show a rapid decrease in the loss value during the initial few epochs and slowly converging afterwards, indicating that the learning and optimization methods of CERNet were effective. The trends of the training and validation datasets were very close, which indicates that CERNet is performing well and is showing very little overfitting, allowing CERNet to be used as an accurate classifier for melanoma.

Figure 4 illustrates the training and validation accuracy and loss values for DEMNet models on the ISIC 2017 dataset. There is a steady convergence of all models in the rules, with minor gaps between training and validation trends, indicating effective learning with controlled overfitting.

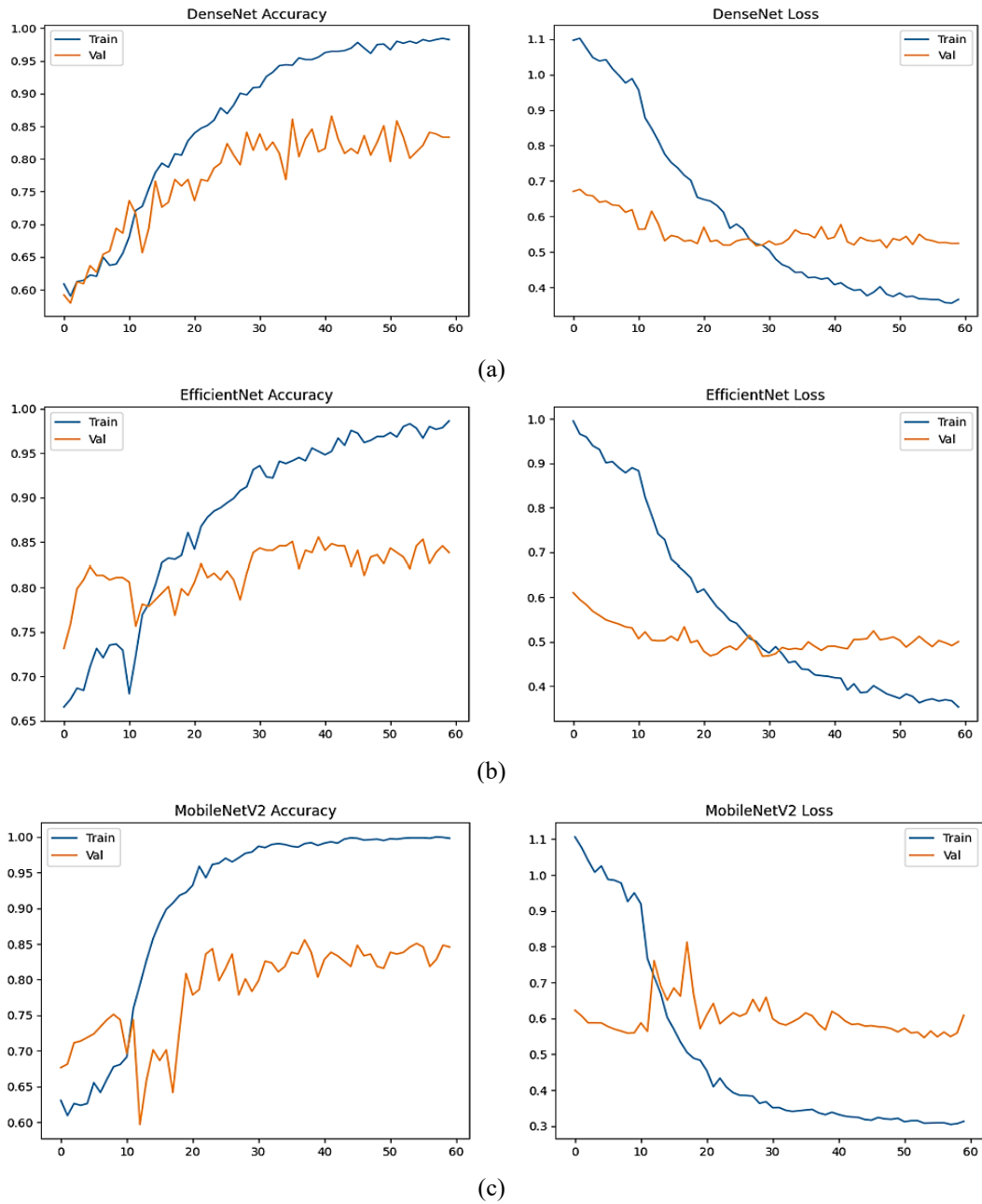


Figure 4. Accuracy and loss trends of the Deep Ensemble Model Network (DEMNet) module during training and validation on ISIC 2017 dataset, (a) DenseNet-121, (b) EfficientNet-B0, and (c) MobileNetV2

6.1 Performance evaluation of individual models

The performance of the individual CNN architectures used in the proposed framework is summarized in Table 3.

6.2 Performance evaluation of ensemble models

For each three-model ensemble, all possible model combinations were evaluated using hard voting, soft voting, and weighted soft voting to identify the optimal ensemble configuration. Hard voting determines the final class through majority voting, whereas soft voting averages the class-probability outputs of the constituent models. In the weighted

voting approach, a PAF strategy was employed, in which models demonstrating superior validation performance, particularly in terms of validation AUC, were assigned greater influence during probability-level fusion. Each ensemble was evaluated on the validation set using five performance metrics: accuracy, precision, recall, F1-score, and AUC, and the combination achieving the best overall validation performance was selected as the final ensemble. During inference, only the selected ensemble was used to classify unseen test images. The probability outputs of the constituent models were fused using the selected voting strategy, and the overall prediction was obtained by identifying the class associated with the highest aggregated probability score.

Table 3. Performance of individual models using HAM10000

Model	Accuracy	Precision	Recall	F1-Score	AUC
DenseNet-121	94%	0.94	0.94	0.94	0.97
EfficientNet-B0	95%	0.95	0.95	0.95	0.97
MobileNetV2	94%	0.94	0.94	0.94	0.96
ConvNeXt-Tiny	91%	0.91	0.90	0.90	0.94
EfficientNetV2-B0	93%	0.93	0.92	0.92	0.96
ResNet50	92%	0.92	0.91	0.91	0.95

Table 4. Performance comparison of ensemble learning approaches

Framework	Model Combination	Voting Strategy	Precision	Recall	F1-Score	Accuracy
DEMNet	DenseNet-121 + EfficientNet-B0	Soft Voting	0.95	0.95	0.95	0.95
	DenseNet-121 + MobileNetV2		0.95	0.95	0.95	0.95
	EfficientNet-B0 + MobileNetV2		0.96	0.96	0.96	0.96
	DenseNet-121 + EfficientNet-B0	Hard Voting	0.95	0.95	0.95	0.95
	DenseNet-121 + MobileNetV2		0.95	0.95	0.95	0.95
	EfficientNet-B0 + MobileNetV2		0.95	0.95	0.95	0.95
	DenseNet-121 + EfficientNet-B0	Weighted Voting	0.95	0.95	0.95	0.95
	DenseNet-121 + MobileNetV2		0.95	0.95	0.95	0.95
	EfficientNet-B0 + MobileNetV2		0.96	0.96	0.96	0.96
	Full Ensemble (DenseNet-121 + EfficientNet-B0 + MobileNetV2)		0.95	0.95	0.95	0.95
	ConvNeXt-Tiny + EfficientNetV2-B0	Hard Voting	0.89	0.88	0.88	0.884
	EfficientNetV2-B0 + ResNet50		0.92	0.92	0.92	0.920
CERNet	ConvNeXt-Tiny + ResNet50	Hard Voting	0.90	0.90	0.90	0.897
	Full Ensemble(ConvNeXt-Tiny + EfficientNetV2-B0 + ResNet50)		0.93	0.93	0.93	0.929
	ConvNeXt-Tiny + EfficientNetV2-B0	Soft Voting	0.92	0.92	0.92	0.917
	EfficientNetV2-B0 + ResNet50		0.93	0.93	0.93	0.932
	ConvNeXt-Tiny + ResNet50		0.93	0.93	0.93	0.933
	Full Ensemble(ConvNeXt-Tiny + EfficientNetV2-B0 + ResNet50)	Weighted Voting	0.94	0.94	0.94	0.939
	ConvNeXt-Tiny + EfficientNetV2-B0		0.91	0.91	0.91	0.913
	EfficientNetV2-B0 + ResNet50		0.93	0.93	0.93	0.934
	ConvNeXt-Tiny + ResNet50	Weighted Voting	0.93	0.93	0.93	0.932
	Full Ensemble (ConvNeXt-Tiny + EfficientNetV2-B0 + ResNet50)		0.94	0.94	0.94	0.937

Note: DEMNet = Deep Ensemble Model Network; CERNet = Convolutional Enhanced Representation Network.

Table 4 presents an exhaustive comparison of the DEMNet and CERNet ensemble configurations using different voting methods to determine the effect of these approaches on melanoma classes (HAM10000 dataset). The ensemble models were evaluated using different voting methods such as hard voting, soft voting, and weighted voting to test how each method affects classification performance for melanomas. Specifically, DEMNet achieved the highest accuracy of 96% from the EfficientNet-B0 and MobileNetV2 ensemble using soft and weighted voting, while the highest accuracy achieved by CERNet is 93.9% from the ensemble of ConvNeXt-Tiny, EfficientNetV2-B0 and ResNet50 using soft voting on the HAM10000 dataset. Results indicate that the use of

heterogeneous CNN-based ensembles improves accuracy and repeatability in classifying melanomas using stored dermoscopic images. The combination of DenseNet-121 and MobileNetV2 is chosen because they reduce feature redundancy and improve the diversity within the ensemble, thereby improving the generalization and robustness of the proposed strategy.

In some cases, the EfficientNet-B0 and MobileNetV2 ensemble slightly outperforms the full ensemble because the two models provide complementary feature representations with less redundancy. Rich multi-scale features are captured by EfficientNet-B0, while MobileNetV2 efficiently learns lightweight discriminative patterns. Thus, a selective dual

ensemble can generalize better than a larger ensemble in some cases. In contrast, expanding the ensemble with additional models can result in redundant or highly correlated predictions, limiting the benefits of increased model diversity. Although performance-based weighting helps balance the contribution of each model, the extra models provide only limited additional information. Consequently, the EfficientNet-MobileNet ensemble demonstrates marginally better performance on certain evaluation metrics while maintaining lower computational complexity.

Figure 5 displays the performance of different ensemble architectures comprising DenseNet-121, EfficientNet-B0, and MobileNetV2 using soft, hard, and weighted voting. The findings demonstrate that ensemble learning strategies give better performance than single-model architectures in terms of accurate melanoma classification. Overall, the ensemble models with different voting methods produce the best results.

Figure 6 displays the performance of different ensemble architectures comprising ConvNeXt-Tiny, EfficientNetV2-B0, and ResNet50 via soft, hard, and weighted voting schemes. The findings demonstrate that ensemble learning strategies outperform single-model approaches in terms of accurate melanoma classification. Overall, the best results were produced by the full ensemble with the weighted voting method, indicating that providing optimized weights to each of the three models provides the greatest level of predictive ability for accurate melanoma classification.

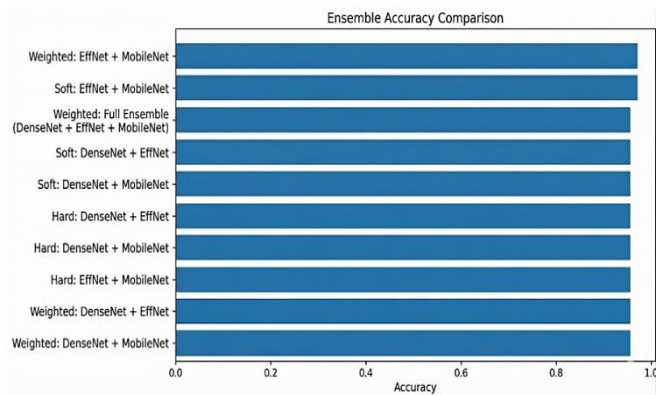


Figure 5. Accuracy comparison of different ensemble model combinations in the Deep Ensemble Model Network (DEMNet)

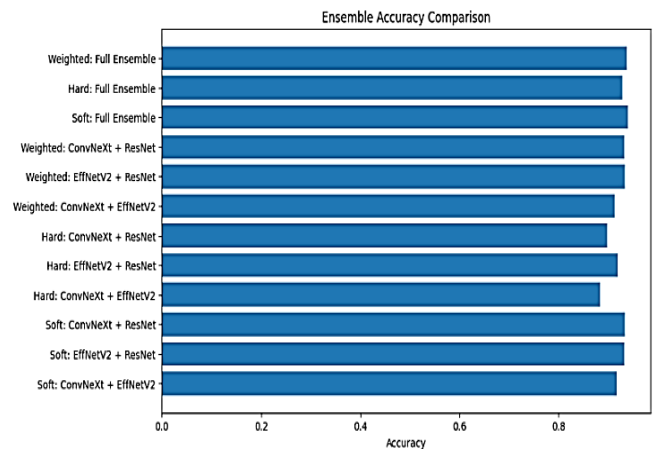


Figure 6. Accuracy comparison of different ensemble model combinations in the Convolutional Enhanced Representation Network (CERNet)

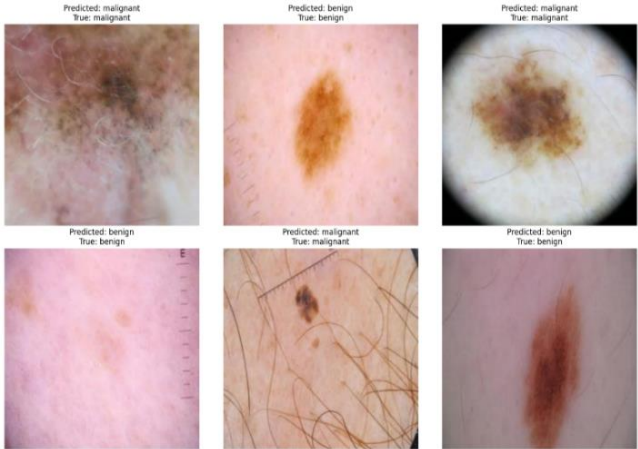


Figure 7. Model predictions and ground truth comparison on the test dataset

Figure 7 presents sample results from the test dataset, illustrating the model’s performance in classifying dermoscopic images as benign or malignant. Each image includes both the predicted label and the true label, demonstrating correct classifications across different lesion types. The results highlight the model’s ability to accurately distinguish between malignant and benign skin lesions, supporting its effectiveness in melanoma detection.

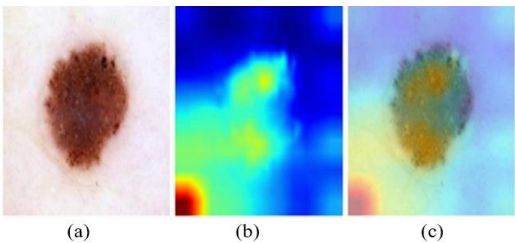


Figure 8. Grad-CAM-based visual interpretation of the proposed DEMNet-CERNet framework: (a) Original dermoscopic image; (b) Grad-CAM heatmap; (c) heatmap overlaid on the original image

Note: Grad-CAM = Gradient-weighted Class Activation Mapping; DEMNet = Deep Ensemble Model Network; CERNet = Convolutional Enhanced Representation Network

Figure 8 illustrates Grad-CAM heatmaps and overlay visualizations highlighting the important regions used by the ensemble model for decision-making. The highlighted areas correspond to critical lesion features, demonstrating the model’s interpretability and focus on relevant regions. Grad-CAM heatmaps were superimposed on the corresponding original dermoscopic images. As shown in Figure 8, the proposed framework consistently highlights the lesion region and its diagnostically relevant characteristics, while assigning marginal attention to the surrounding healthy skin. This indicates that the model learns clinically meaningful features rather than relying on background artifacts. The visualization supports the reliability of the proposed dual-ensemble framework and provides evidence that the classification decisions are based on relevant lesion characteristics.

In Table 5, the authors compare the effectiveness of the new DEMNet and CERNet models with existing state-of-the-art dermoscopic image classification methods using several benchmark datasets (HAM10000, ISIC 2016 and ISIC 2017). The results show that the DEMNet and CERNet models

perform better than previously reported works for the HAM10000 dataset but have an equivalent level of cross-dataset evaluation performance. These results provide strong evidence that the proposed ensemble strategy enhances classification accuracy and feature representation, as well as the generalization ability across dermoscopic images from multiple datasets, thereby outperforming previous models.

The improved performance of the proposed framework is attributed to the combined effect of image preprocessing, complementary feature learning, and the proposed PAF strategy. Image preprocessing improves the quality and

consistency of dermoscopic images, while the selected CNN architectures learn complementary feature representations that enhance lesion characterization. The PAF strategy further improves prediction by assigning greater importance to models with superior validation performance during probability-level fusion. Together, these components improve the robustness, generalization, and classification performance of the proposed framework. Since an ablation study was not conducted, the contribution of each component cannot be quantified independently.

Table 5. Comparative analysis with existing skin lesion classification methods on dermoscopic datasets

Method	Dataset	Accuracy	Precision	Recall	F1-Score
Kaur et al. [26]	ISIC 2016	0.81	0.81	0.81	0.81
Kaur et al. [26]	ISIC 2017	0.88	0.78	0.87	0.78
Hoang et al. [27]	HAM10000	0.84	0.75	---	0.72
DEMNet (Proposed)	HAM10000	0.95	0.95	0.95	0.95
CERNet (Proposed)	HAM10000	0.93	0.93	0.93	0.93
DEMNet (Proposed)	ISIC 2016	0.82	0.83	0.81	0.82
CERNet (Proposed)	ISIC 2017	0.82	0.84	0.81	0.82

Note: DEMNet = Deep Ensemble Model Network; CERNet = Convolutional Enhanced Representation Network.

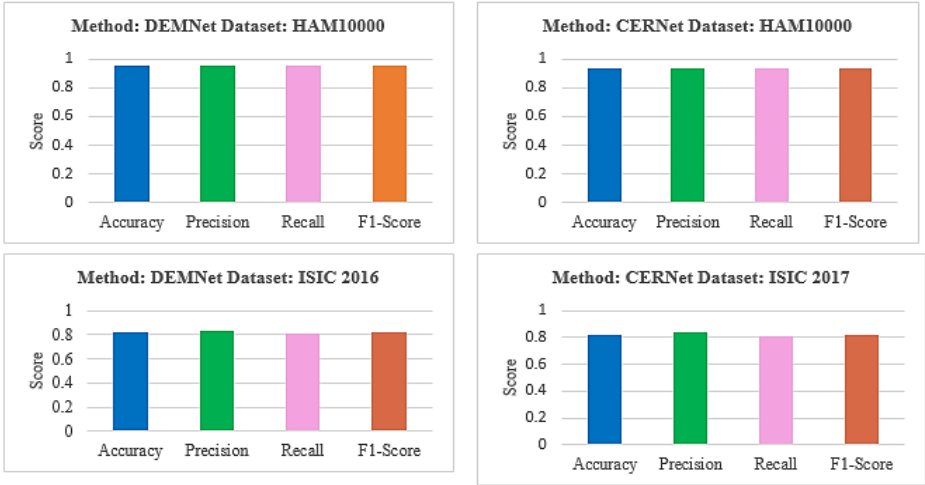


Figure 9. Comparative analysis of classification performance for proposed models across ISIC and HAM10000 datasets

Figure 9 illustrates that the proposed models demonstrate robust performance across multiple benchmark datasets, with the DEMNet architecture achieving the highest results on the HAM10000 dataset, with an accuracy of 0.95. While the DEMNet and CERNet models maintained stable performance across the ISIC 2016 and ISIC 2017 datasets with a consistent accuracy of 0.82, a significant upward trend in evaluation metrics (including precision, recall, and F1-Score) is observed when transitioning to the HAM10000 dataset, where both CERNet and DEMNet consistently exceeded the 0.93 threshold. This indicates that the proposed methodologies are highly effective at feature extraction and classification, particularly as dataset complexity and volume increase.

7. CONCLUSION AND FUTURE WORK

A new dual-ensemble DL framework (DEMNet–CERNet) has been proposed for skin lesion classification that integrates image preprocessing with complementary DL architectures to improve performance. The DEMNet uses multiple levels of hierarchical features to provide improved feature

representation and reduce model bias. The complementary learning of deep features that creates a diverse set of classification layers improves the robustness of classification in CERNet. The proposed dual-ensemble framework has been evaluated on multiple benchmark dermoscopy datasets (HAM10000, ISIC 2016, and ISIC 2017) to demonstrate that the proposed ensemble strategy outperforms the individual models and attains competitive performance across multiple performance metrics. The adaptive weighted fusion strategy exploits the complementary strengths of selected architectures, resulting in improvements in classification accuracy, robustness, and model generalization. The important finding of this research work is that a wisely selected dual ensemble, combined with proper image preprocessing and a suitable weighting scheme, can achieve improved classification accuracy. Additionally, the Grad-CAM visualizations ensure that the proposed framework focuses on clinically relevant lesion regions, which enhances the interpretability and reliability of its predictions. The findings of this research work establish that the proposed framework can act as an effective computer-aided diagnostic system for early skin lesion detection.

Future research could further enhance the proposed framework in several ways. One improvement would be the use of advanced transformer-based architectures (such as Vision Transformers) to improve the classification performance. Future research could also investigate the incorporation of the proposed framework with federated learning for collaborative melanoma classification across hospitals and dermatology centers, enabling the hospitals to train the model locally by sharing only the model updates. This approach would enable the proposed framework to handle diverse skin lesion datasets without compromising patient confidentiality.

REFERENCES

- [1] Tschandl, P., Rosendahl, C., Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 5: 180161. <https://doi.org/10.1038/sdata.2018.161>
- [2] Esteva, A., Kuprel, B., Novoa, R.A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639): 115-118. <https://doi.org/10.1038/nature21056>
- [3] Brinker, T.J., Hekler, A., Enk, A.H., et al. (2019). Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermatoscopic melanoma image classification task. *European Journal of Cancer*, 113: 47-54. <https://doi.org/10.1016/j.ejca.2019.04.001>
- [4] Kassani, S.H., Kassani, P.H. (2019). A comparative study of deep learning architectures on melanoma detection. *Tissue and Cell*, 58: 76-83. <https://doi.org/10.1016/j.tice.2019.04.009>
- [5] Sharma, A.K., Tiwari, S., Aggarwal, G., Goenka, N., Kumar, A., Chakrabarti, P. (2022). Dermatologist-level classification of skin cancer using cascaded ensembling of convolutional neural network and handcrafted features based deep neural network. *IEEE Access*, 10: 17920-17932. <https://doi.org/10.1109/ACCESS.2022.3149824>
- [6] Kausar, N., Hameed, A., Sattar, M., et al. (2021). Multiclass skin cancer classification using ensemble of fine-tuned deep learning models. *Applied Sciences*, 11(22): 10593. <https://doi.org/10.3390/app112210593>
- [7] Qureshi, A.S., Roos, T. (2023). Transfer learning with ensembles of deep neural networks for skin cancer detection in imbalanced data sets. *Neural Processing Letters*, 55: 4461-4479. <https://doi.org/10.1007/s11063-022-11049-4>
- [8] Gessert, N., Nielsen, M., Shaikh, M., Werner, R., Schlaefer, A. (2020). Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data. *MethodsX*, 7: 100864. <https://doi.org/10.1016/j.mex.2020.100864>
- [9] Tsai, W.X., Li, Y.C., Lin, C.H. (2023). Skin lesion classification based on multi-model ensemble with generated levels-of-detail images. *Biomedical Signal Processing and Control*, 85: 105068. <https://doi.org/10.1016/j.bspc.2023.105068>
- [10] Bian, J.X., Zhang, S., Wang, S.Q., Zhang, J.R., Guo, J.C. (2021). Skin lesion classification by multi-view filtered transfer learning. *IEEE Access*, 9: 66052-66061. <https://doi.org/10.1109/ACCESS.2021.3076533>
- [11] Juan, C.K., Su, Y.H., Wu, C.Y., et al. (2023). Deep convolutional neural network with fusion strategy for skin cancer recognition: Model development and validation. *Scientific Reports*, 13: 17087. <https://doi.org/10.1038/s41598-023-42693-y>
- [12] Nawaz, M., Mehmood, Z., Nazir, T., et al. (2022). Skin cancer detection from dermoscopic images using deep learning and fuzzy k-means clustering. *Microscopy Research and Technique*, 85(1): 339-351. <https://doi.org/10.1002/jemt.23908>
- [13] Toprak, A.N., Aruk, I. (2024). A hybrid convolutional neural network model for the classification of multi-class skin cancer. *International Journal of Imaging Systems and Technology*, 34(5): e23180. <https://doi.org/10.1002/ima.23180>
- [14] Venugopal, V., Raj, N.I., Nath, M.K., Stephen, N. (2023). A deep neural network using modified EfficientNet for skin cancer detection in dermoscopic images. *Decision Analytics Journal*, 8: 100278. <https://doi.org/10.1016/j.dajour.2023.100278>
- [15] Shakya, M., Patel, R., Joshi, S. (2025). A comprehensive analysis of deep learning and transfer learning techniques for skin cancer classification. *Scientific reports*, 15(1): 4633. <https://doi.org/s41598-024-82241-w>
- [16] Manole, I., Butacu, A.I., Bejan, R.N., Tiplica, G.S. (2024). Enhancing dermatological diagnostics with efficientnet: A deep learning approach. *Bioengineering*, 11(8): 810. <https://doi.org/10.3390/bioengineering11080810>
- [17] Vieira, J., Mendonça, F., Morgado-Dias, F. (2025). Deep learning approaches for skin lesion detection. *Electronics*, 14(14): 2785. <https://doi.org/10.3390/electronics14142785>
- [18] Cherukuri, S., Vara Prasad, S.D. (2025). EnsembleSkinNet: A transfer learning-based framework for efficient skin cancer detection with explainable AI integration. *Frontiers in Oncology*, 15: 1699960. <https://doi.org/10.3389/fonc.2025.1699960>
- [19] V, U., Shivaram, J.M., Guggari, S., Okoye, K. (2025). Ensemble method of pre-trained models for classification of skin lesion images. *Applied Sciences*, 15(24): 13083. <https://doi.org/10.3390/app152413083>
- [20] Chand, H.V., Sabharwal, S., Jiang, W., Islam, S.M., Cengiz, K. (2026). A stacking-based deep learning ensemble for multi-class skin lesion classification. *Discover Computing*, 29(1): 417. <https://doi.org/10.1007/s10791-026-10288-6>
- [21] Sethanan, K., Pitakaso, R., Srichok, T., et al. (2023). Double AMIS-ensemble deep learning for skin cancer classification. *Expert Systems with Applications*, 234: 121047. <https://doi.org/10.1016/j.eswa.2023.121047>
- [22] Chatterjee, S., Gil, J.M., Byun, Y.C. (2024). Early detection of multiclass skin lesions using transfer learning-based IncepX-ensemble model. *IEEE Access*, 12: 113677-113693. <https://doi.org/10.1109/ACCESS.2024.3432904>
- [23] Chanda, D., Onim, M.S.H., Nyeem, H., Ovi, T.B., Naba, S.S. (2024). DCENSnet: A new deep convolutional ensemble network for skin cancer classification. *Biomedical Signal Processing and Control*, 89: 105757. <https://doi.org/10.1016/j.bspc.2023.105757>
- [24] Dakhli, R., Barhoumi, W. (2024). Disagreement serves an explainable ensemble model based on Dempster-Shafer evidence-fusion for an improved skin lesion classification. *Biomedical Signal Processing and*

- Control, 98: 106761.
<https://doi.org/10.1016/j.bspc.2024.106761>
- [25] Codella, N.C., Gutman, D., Celebi, M.E., et al. (2018). Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). In 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 168-172.
<https://doi.org/10.1109/ISBI.2018.8363547>
- [26] Kaur, R., GholamHosseini, H., Sinha, R., Lindén, M. (2022). Melanoma classification using a novel deep convolutional neural network with dermoscopic images. *Sensors*, 22(3): 1134. <https://doi.org/10.3390/s22031134>
- [27] Hoang, L., Lee, S.H., Lee, E.J., Kwon, K.R. (2022). Multiclass skin lesion classification using a novel lightweight deep learning framework for smart healthcare. *Applied Sciences*, 12(5): 2677.
<https://doi.org/10.3390/app12052677>