COMPUTING AND ALGORITHM INSIGHT
Volume 4, Issue 1, June 2026 | pp. 1-22
https://doi.org/10.63208/21018-239

OPEN ACCESS
Check for updates

ENHANCED ROBUSTNESS IN AUTOMATED SKIN CANCER DETECTION VIA A CNN-BASED ENSEMBLE CLASSIFIER

Li Wei[1] ✉, Zhang Ming[1], and Wang Xiaoyu[2]

School of Computer Science and Technology, Jiangsu University, Zhenjiang, China[1]; School of Information and Electrical Engineering, Shaoyang University, Shaoyang, China[2]

ABSTRACT

Malignant skin lesions, commonly referred to as skin cancer, represent a leading cause of global mortality. The integration of advanced deep learning methodologies facilitates the early diagnosis of these lesions, thereby offering a vital pathway to decrease associated mortality rates. For the automated categorization of dermoscopic images, deep convolutional neural networks (CNNs) have consistently exhibited performance superior to that of traditional machine learning algorithms. The current research focuses on the development of a highly precise ensemble classifier built upon foundational CNN architectures for the purpose of classifying skin lesions. To ensure the broad generalizability of the proposed model, strategic enhancements in diversity were implemented to bolster both systemic resilience and accuracy. Specifically, this necessary variance was introduced at both the dataset and the algorithmic classifier levels. The comprehensive ISIC 2018 challenge dataset, encompassing more than 13,000 dermoscopic images, was utilized across the training, validation, and testing phases. Experimental outcomes indicate that the developed ensemble framework attained an accuracy rate of 90.15%, successfully surpassing baseline CNN configurations and comparable contemporary classifiers in terms of overall accuracy and robust performance.

Keywords: *Deep Learning; Convolutional Neural Networks; Ensemble Classification; Dermoscopy; Skin Lesion Diagnosis*

CORRESPONDING AUTHOR:
Li Wei
School of Computer Science and Technology, Jiangsu University, Zhenjiang, China
contact: *liwei@ujs.edu.cn*

ARTICLE HISTORY
Received : December 15, 2025
Final Revised : May 8, 2026
Accepted : June 2, 2026
Published : June 30, 2026

© 2026 The Author(s) | Published by Integra Academic Press | This article is an open access article and distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license.

1. | INTRODUCTION

Cutaneous malignancies represent a critical health threat characterized by substantial global mortality rates, with authoritative international studies classifying skin cancer among the most lethal and prevalent oncological diseases [1]. Data from the World Health Organization highlights a continuous rise in incidence over recent decades, with an estimated global occurrence of 2 to 3 million non-melanoma cases alongside 132,000 melanoma instances annually [2]. Furthermore, the National Cancer Institute (NCI) identifies melanoma as the most frequently diagnosed cancer in the young adult demographic, specifically those aged 25 to 29 [3]. Within the United States, approximately 9% of the population is afflicted by this specific malignancy [1]. In Canada, it currently ranks as the seventh most prevalent cancer, and notably, the cumulative cases of skin cancer over the last three decades have surpassed the combined total of all other cancer types [4]. The situation is particularly acute in Australia, where skin lesions account for 10% of all malignancies, demonstrating an incidence rate quadruple that of the US, UK, and Canada [5]. Driven by the persistent global surge in cases, corresponding high mortality, and significant economic healthcare burdens, achieving early detection remains a critical public health imperative. Prompt identification and subsequent intervention mitigate the probability of metastasis to vital organs, such as the liver, lungs, or brain, while simultaneously averting the need for invasive and expensive procedures like chemotherapy, radiotherapy, and extensive biopsies. Crucially, early-stage therapeutic management is highly effective, potentially elevating patient survival rates to 95% [6], [7], [8]. In clinical practice, the diagnostic process heavily relies on the visual appraisal of dermoscopic imagery by dermatologists; a methodology inherently constrained by the practitioner's level of expertise, specific training, and visual acuity [6], [9]. Consequently, the integration of automated computational frameworks offers substantial support to medical professionals in refining diagnostic accuracy [8]. Historically, the development of computer-aided diagnosis (CADx) systems primarily leveraged conventional computer vision methodologies, encompassing phases such as segmentation, manual feature extraction, selection, and subsequent classification [3], [5], [9], [10], [11], [12], [13], [14]. While these traditional systems yielded adequate results, their scope was largely restricted to binary categorization (benign versus malignant). The execution of multi-class categorization persisted as a formidable challenge due to substantial intra-class variances and inter-class similarities. More recently, deep learning paradigms have gained prominence, demonstrably outperforming classic image processing techniques in complex classification tasks [15]. Specifically, deep Convolutional Neural Networks (CNNs) have exhibited exceptional efficacy in the automated identification and diagnosis of dermal lesions when contrasted with traditional machine learning approaches [2], [16]. Nevertheless, the profound dependency of deep learning models on vast repositories of meticulously annotated training data remains a significant logistical hurdle [15].

To optimize diagnostic capabilities within this domain, numerous contemporary approaches advocate for the deployment of deep neural network ensembles, where the maximization of model diversity is fundamental to attaining superior classification accuracy. The current investigation amplifies this diversity initially at the data stratum through the implementation of bootstrapping methodologies. By executing successive training iterations, the resultant models are exposed to distinct, varied subsets of the foundational data. Because

subsets of these models inevitably manifest comparable classification outputs and high inter-model agreement due to their underlying training distributions, optimizing diversity at the classifier level emerges as a critical necessity; this precise optimization constitutes the primary novelty of our study. Our framework not only integrates a variety of heterogeneous convolutional architectures to foster this classifier-level variance, but it also introduces an innovative filtering mechanism utilizing the Cohen's Kappa score to systematically eliminate highly correlated, redundant classifiers. Empirical evaluations confirm that the strategic pruning of these analogous classifiers via the proposed metric substantially elevates the overall precision of the ensemble, thereby validating the methodological efficacy of this approach.

2. | RELATED WORKS

Nyiri and Kiss [17] explored advanced techniques for creating ensembles of deep neural networks, utilizing diverse hyperparameters and distinct data preprocessing strategies for the classification of skin lesions. Their study, employing both the ISIC 2017 and ISIC 2018 datasets, demonstrated that ensemble methods are advantageous not only for aggregating different machine learning architectures but also for synthesizing varied hyperparameter configurations. Harangi [15] examined the potential for enhancing the accuracy of deep convolutional neural networks through an ensemble approach designed to categorize dermoscopic images into three specific classes: melanoma, nevus, and seborrheic keratosis. This was achieved by integrating the outputs from the classification layers of four distinct deep neural network architectures, with the ultimate classification determined via the weighted output of these robust foundational CNNs. Similarly, Goyal and Rajapakse [18] constructed an ensemble model incorporating two specific DNN architectures Inception-v4 and InceptionResnet-v2 which were synthesized using a 1×1 convolution within a meta-learning fusion layer.

Nanni et al. [19] introduced an ensemble of descriptors specifically for melanoma classification, evaluating its efficacy against the Melanoma Challenge 2018. Their system achieved notable discriminative capability by merging multiple descriptors derived from varied CNN ensemble architectures and hyperparameter settings. Furthermore, they utilized pre-trained CNNs as feature extractors, where the output of a specific layer served as an image descriptor, subsequently used to train a suite of Support Vector Machines (SVMs). Guo and Ashour [20] developed a multiple convolution neural network (MCNN) framework to classify diverse skin lesions within dermoscopic images. Utilizing an additive sample learning strategy, they independently trained several models on the ISIC 2018 training and validation sets, subsequently evaluating performance through receiver operating characteristic (ROC) curves and the corresponding area under the curve (AUC). Xie et al. [21] proposed an end-to-end multi-level deep ensemble (MLDE) model utilizing four pre-trained ResNet-50 networks to process multi-scale lesion data, integrated via an adaptive weighting mechanism.

Gessert et al. [22] submitted an algorithm to the ISIC 2018 challenge designed to classify seven lesion categories across 10,015 images. They implemented a CNN ensemble incorporating architectures such as DenseNet, SENet, and ResNeXt. Addressing significant class imbalance, they prioritized strategies like loss weighting and balanced batch sampling, ultimately employing meta-learning techniques for final predictive outcomes. Majtner et al. [23] detailed an automated diagnostic system that applied transfer learning to fine-tune VGG16 [24] and GoogLeNet-based [25] architectures. A primary contribution of their methodology

was the implementation of robust image augmentation and color normalization during preprocessing, which was evaluated against Task 3 of the ISIC 2018 dataset. Kitada and Iyatomi [26] engineered a highly effective image recognition system utilizing multiple pre-trained Squeeze-and-Excitation Networks (SENet), operating within a semi-supervised learning framework with specific data augmentation strategies. Their approach yielded a balanced accuracy of 87.2% on the official ISIC 2018 validation set.

Serte and Demirel [2] developed a classification method predicated on decomposing input images into seven directional sub-bands using a Gabor-based approach. These sub-bands, alongside the original image, were processed by eight parallel CNNs, with the final lesion classification determined via a Summation Rule for decision fusion. Another study [7] assessed the efficacy of the EfficientNet family in lesion classification, fine-tuning architectures B0 through B7 on the HAM10000 dataset [27]. The EfficientNet-B0 model achieved optimal performance, registering 87.9% accuracy, 88% precision and recall, and an AUC of 97.5%; however, comparative analysis with contemporary methodologies was limited. Research detailed in [28] employed a two-step procedure: initial lesion segmentation using an optimized region-growing algorithm, followed by classification via an autoencoder model. While this method achieved 94.2% accuracy on the PH2 dataset [29], its reliance on segmentation quality and lack of evaluation on larger datasets like HAM10000 remain notable drawbacks.

Further research [30] proposed a deep CNN incorporating a multi-scale attention mechanism and loss weighting to counteract class imbalance, achieving 91.6% accuracy on the HAM10000 dataset. Another study [31] demonstrated that integrating a soft attention mechanism with foundational CNN backbones (such as VGG, ResNet, Inception, and DenseNet) improved precision and recall by 4.7% and 3.8%, respectively. A multi-modal strategy was proposed in [32], processing visual and textual lesion features concurrently and merging them via soft attention, resulting in 83% accuracy on the 7-point criteria dataset. A four-phase methodology outlined in [33] encompassing median filter noise reduction, Otsu's segmentation, feature extraction via 2D DWT and GCLM, and final DNN classificationyielded 84.5% accuracy on the ISIC 2017 test set. This test set accuracy was subsequently improved by 10.3% in study [34] through the deployment of a vision transformer featuring 24 hidden layers as the primary classification backbone.

While the majority of research in this domain focuses on developing ensemble classifiers for skin lesions, the inherent diversity among the constituent models frequently remains inadequate. As emphasized by Kuncheva [35], robust diversity is critical to effective ensembling. Provided sufficient diversity exists, highly accurate ensemble classifiers can be constructed using base models that exhibit only moderate average accuracies (approximately 60%–70%). The utilization of such base classifiers offers a significant advantage by enhancing overall model robustness, thereby mitigating the risk of overfitting a prevalent issue when tuning deep CNNs with extensive parameters on limited datasets. Consequently, the present study endeavors to construct a novel CNN-based ensemble classifier characterized by substantial diversity at both the data and classifier levels to optimize skin lesion classification.

3. | MATERIALS AND METHODS

3.1 Dataset

To train, validate, and evaluate the proposed diagnostic framework, the ISIC 2018 repository was employed. This comprehensive archive encompasses more than 13,000

dermoscopic images sourced from prominent international clinical institutions, captured using a diverse array of acquisition equipment at each respective site. The extensive, multinational participation in compiling this visual data was intentionally structured to guarantee a highly representative and clinically applicable sample [27], [36], [37]. Furthermore, the archive systematically classifies the data into seven distinct dermatological pathologies: Melanoma, Melanocytic nevus, Basal cell carcinoma, Actinic keratosis / Bowen's disease (intraepithelial carcinoma), Benign keratosis (encompassing solar lentigo, seborrheic keratosis, and lichen planus-like keratosis), Dermatofibroma, and Vascular lesions [36]. Representative visual specimens of these diverse skin anomalies are illustrated in Fig. 1.

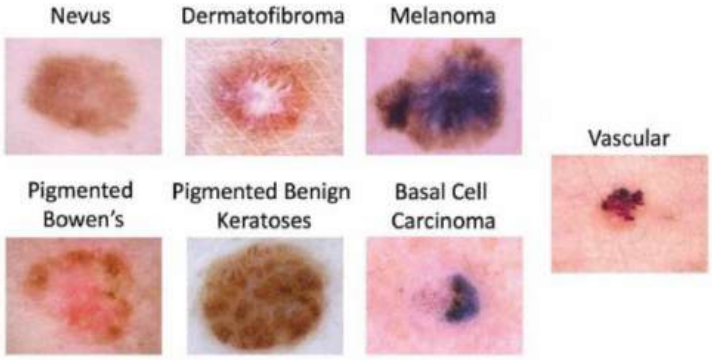


Figure 1. Representative samples of the seven skin lesion categories.

The ISIC 2018 lesion imagery originates from the HAM10000 dataset and was captured utilizing diverse dermatoscopic instruments. The class distribution of these pathological states reflects an adjusted clinical reality; although benign lesions consistently outnumber their malignant counterparts, malignancies are intentionally over-represented within the collection [36]. The specific quantitative allocation of images across each diagnostic category is delineated in Fig. 2. Furthermore, owing to the prevalent adoption of the HAM10000 dataset as a standard benchmark within this research domain, it is similarly employed in the present study to evaluate and report the efficacy of the proposed methodology.

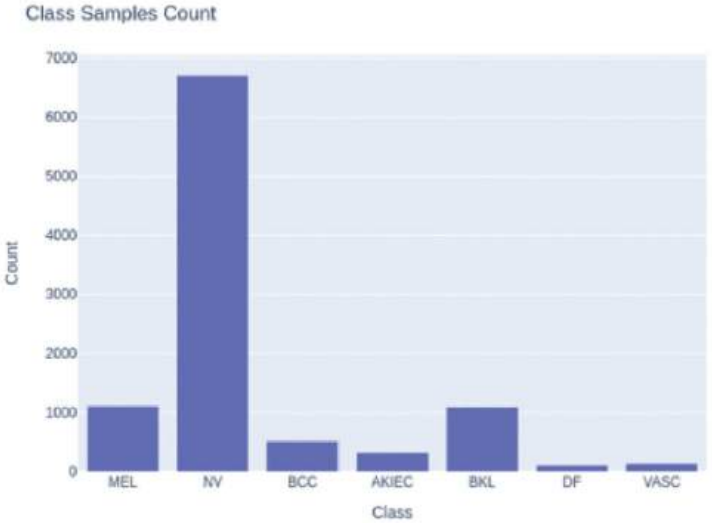


Figure 2. Distribution of the seven skin lesion categories within the ISIC 2018 dataset.

3.2 Data Pre-Processing and Augmentation

For the training, validation, and evaluation of the proposed methodology, the original data partitioning established by the ISIC 2018 dataset was preserved, allocating 10,015 images to the training set, 193 to the validation set, and 1,512 to the testing set. A comprehensive enumeration of the sample distribution across each diagnostic class within these subsets is detailed in Table 1. To systematically counteract the inherent class imbalance, an equalization strategy was enforced during batch construction in the training phase, guaranteeing uniform sample representation across all categories. Additionally, the implementation of stochastic data augmentations substantially magnified the effective volume of training instances exposed to the base classifiers. To rigorously assess the framework's performance on the HAM10000 dataset, a 5-fold cross-validation protocol was adopted. In this configuration, each randomized fold designated 80% of the corpus for training and the remaining 20% for testing, with the ultimate performance outcomes computed as the statistical average across all five distinct folds.

Table 1. Sample sizes categorized by class and data split within the ISIC 2018

Class	Train	Validation	Test
MEL	1113	21	171
NV	6705	123	909
BCC	514	15	93
AKIEC	327	8	43
BKL	1099	22	217
DF	115	1	44
VASC	142	3	35

The preprocessing protocol commences with image normalization, executed by subtracting the ImageNet dataset's mean value from each image and subsequently dividing by its standard deviation. Furthermore, to eliminate peripheral, non-informative areas within the dermoscopic scans, a center-cropping procedure is implemented. This cropping is standardized to dimensions of 240×240 pixels, thereby establishing a uniform input size of 240×240 for the neural networks. Crucially, these preprocessing operations are uniformly applied across all images within the training, validation, and testing subsets prior to their introduction into the network architecture.

Data augmentation constitutes a critical component of the machine learning pipeline, exerting a profound influence on overall model efficacy. The strategic deployment of augmentation techniques serves to artificially expand and diversify the dataset, thereby mitigating the risk of overfitting, particularly when training on limited data volumes. Throughout the training phase, a diverse array of augmentation strategies was employed to synthesize additional training instances, a measure especially beneficial for underrepresented classes. These techniques encompassed vertical and horizontal flipping, random rotations ranging from -30 to 30 degrees, and color jittering, which introduced random fluctuations in brightness and contrast up to a maximum variance of 0.1. Additionally, random affine transformations were integrated, synergizing random rotations (-180 to 180 degrees) with spatial translations (0.1 along both axes) and scaling adjustments (spanning 0.7 to 1.3). This comprehensive augmentation regimen was designed to significantly enhance the model's resilience and generalizability against diverse spatial and visual transformations.

3.3 Methods

The majority of deep learning frameworks operate under a supervised paradigm, wherein neural networks discern system dynamics through training on meticulously labeled datasets that pair specific inputs with corresponding outputs. Convolutional Neural Networks (CNNs) represent a specialized category of artificial neural networks designed to autonomously learn hierarchical representations directly from raw data. This is achieved via a sequence of structural components, including convolutional, pooling, and fully connected layers, facilitating a process often referred to as unsupervised feature extraction. Consequently, the deployment of such architectures effectively obviates the necessity for manual, expert-driven feature engineering. Moreover, the integration of ensemble learning techniques can substantially elevate classification efficacy, a critical requirement in complex domains like skin lesion categorization where exceptional accuracy is mandated. The cornerstone of robust ensemble prediction lies in model diversity; amalgamating disparate models permits the capitalization on varied architectural strengths while simultaneously mitigating individual algorithmic vulnerabilities. The subsequent section elaborates on the specific architectural and operational details of the proposed diagnostic framework.

Figure 3 presents the schematic block diagram detailing the proposed methodology, including a comprehensive illustration of the training protocol for the ensemble classifiers. During this phase, rather than relying on a solitary CNN, an array of distinct classifiers is trained concurrently to construct a highly precise composite model. The core objective of this strategy is to amplify the overall predictive accuracy by synergistically fusing these foundational CNNs. To expedite the computational training phase, transfer learning paradigms [38] were implemented. Accordingly, several constituent classifiers were initialized utilizing weights from pre-trained CNN architectures, specifically SEResNet, DenseNet, and EfficientNet. Within the transfer learning framework, a model previously optimized on a massive, generalized dataset functions as a robust feature extractor. Subsequently, only the terminal layers and the classification head are subjected to weight updates during the fine-tuning phase on the targeted, domain-specific dataset. Given that foundational morphological filters are universally applicable across diverse image modalities, freezing these initial convolutional layers circumvents the redundant recalculation of preliminary weights. Furthermore, transfer learning proves exceptionally advantageous when operating with constrained datasets. A restricted volume of training samples significantly exacerbates the risk of overfitting in deep neural networks due to their vast parameter spaces. By converting a substantial portion of the network into a non-trainable state, transfer learning effectively curtails the number of optimizable parameters, thereby successfully mitigating the overfitting phenomenon.

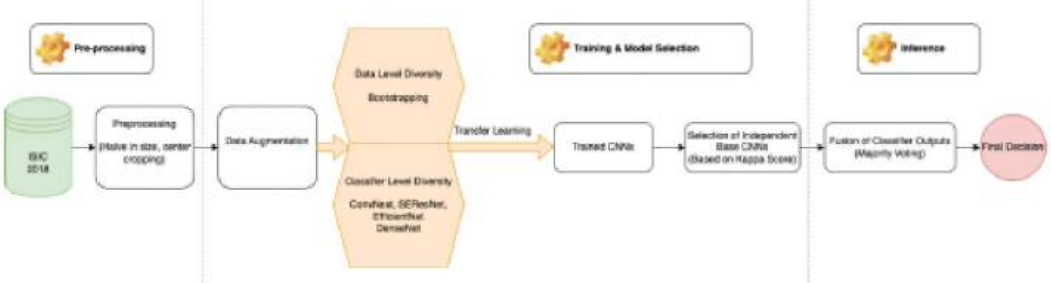


Figure 3. Diagram of the proposed method.

Throughout the training phase, the principal focus remained on the development of highly diverse CNNs. As delineated in Fig. 4, this diversity can be actualized across four distinct strata: the data level, the feature level, the classifier level, and the combiner level [35]. The core objective of this study is to systematically generate variance specifically at the data and classifier levels, strategies for which are explicated in the subsequent subsections. Importantly, owing to the unique structural components of convolutional neural networks relative to conventional machine learning techniques, establishing diversity at the data level inherently facilitates a corresponding diversity at the feature level.

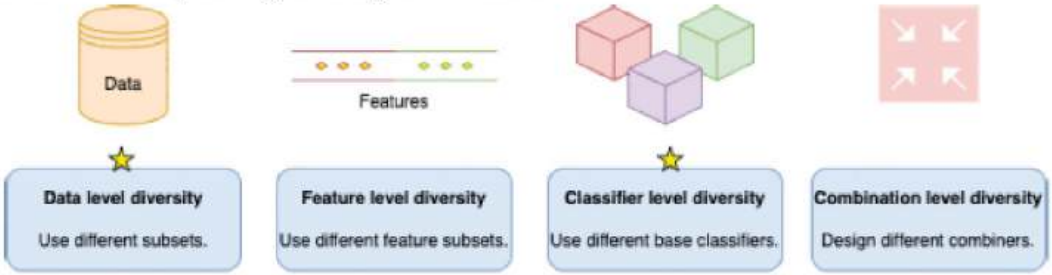


Figure 4. Overview of diversity hierarchies. The targeted data and classifier levels are denoted by stars.

Following the training phase, highly correlated CNNs are pruned utilizing the Kappa score to optimize ensemble diversity, computational efficiency, and spatial complexity. Subsequently, the predictions of the retained classifiers are aggregated via a majority voting mechanism to determine the final classification on the test set. The specific methodologies governing these selection and fusion processes are detailed in the subsequent subsections.

Fostering data-level diversity produces a heterogeneous ensemble, as training inherently unstable models like neural networks on varied subsets generates independent classifiers. To construct these distinct training sets, we employed bootstrapping. A primary challenge during this phase was the severe class imbalance, as illustrated in Fig. 2. To mitigate this and prevent network bias toward majority categories, we applied random equal-frequency sampling across all classes. Each base CNN was subsequently trained on a unique bootstrap generated via random sampling with replacement from the original dataset, with the resulting sample configurations depicted in Fig. 5.

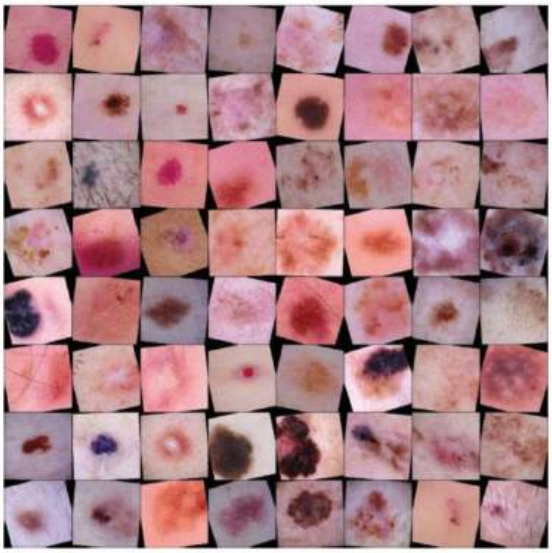


Figure 5. Examples of augmented training data for an individual CNN, uniformly sampled with replacement from each diagnostic category.

To construct a robust and highly accurate ensemble, it is essential to train competent base classifiers that introduce structural diversity at the classifier level. We evaluated various architectures, including VGG16, Inception, Xception, ResNet10, ResNet50, InceptionResNet, DenseNet, and EfficientNet B0-B7. Among these, ConvNext, SENet, DenseNet, and EfficientNet-B0 demonstrated superior accuracy.

Briefly, ConvNext [39] advances prior ResNet-style designs by incorporating an expanded array of building layers (such as layer normalization and depthwise separable convolutions), LayerScale initialization, extreme stochastic depth for improved regularization, and high-resolution feature maps that bypass traditional bottleneck compression. Similarly, SENet [40] enhances the foundational ResNet architecture through the integration of Squeeze-and-Excitation blocks. This attention-based mechanism facilitates adaptive feature recalibration, allowing the model to dynamically emphasize salient information and substantially improve overall network performance, as illustrated in Fig. 6.

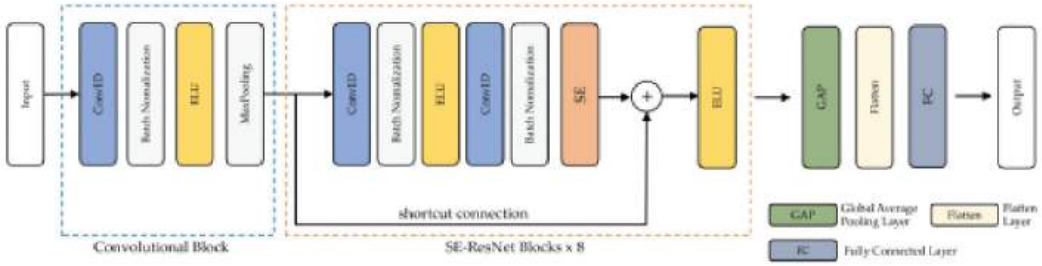


Figure 6. Illustration of the SEResNet architecture [41]. For 2D applications, the depicted Conv1D operations are replaced by Conv2D layers.

EfficientNet [42]: This architecture achieves computational efficiency by either refining a baseline model or algorithmically discovering an optimal structure. While traditional compression methods including network pruning, data quantization, tensor decomposition, and knowledge distillation modify existing models, advanced techniques like Neural Architecture Search (NAS) programmatically identify highly efficient configurations. These primary design methodologies are illustrated in Fig. 7.

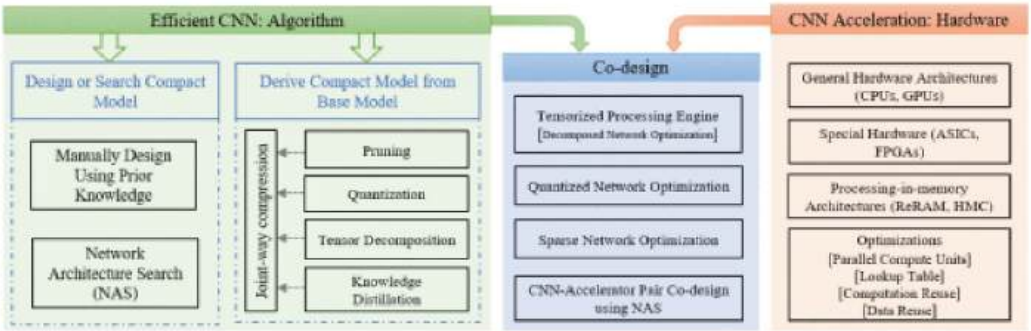


Figure 7. Efficient CNN design approaches

DenseNet [43]: Densely Connected Convolutional Networks (DenseNet) represent a structural deviation from conventional CNNs by abandoning strictly sequential information flow. Instead, DenseNet employs dense connectivity, ensuring that each layer directly receives feature maps from all preceding layers. This architecture promotes extensive feature reuse, mitigates the vanishing gradient problem, and facilitates robust feature propagation, ultimately

resulting in a highly parameter-efficient model. The foundational structure of DenseNet is depicted in Fig. 8.

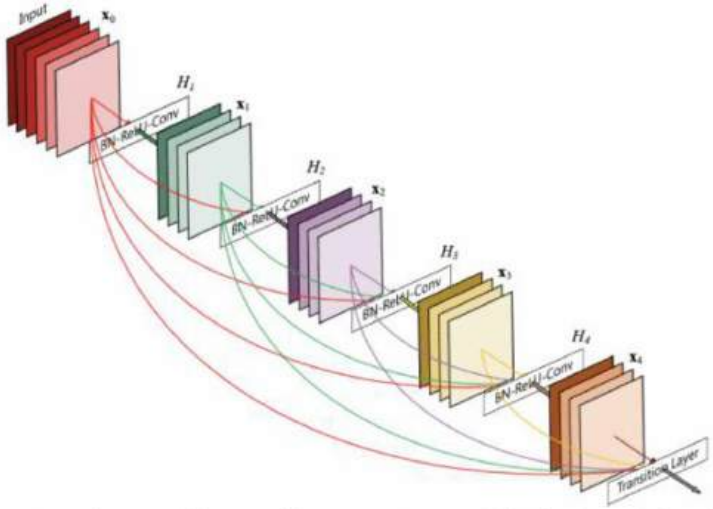


Figure 8. Architectural layout of the DenseNet model [43], adapted from [44].

To optimize ensemble diversity and reduce computational complexity, highly correlated classifiers were systematically excluded utilizing the Cohen's Kappa score, calculated as:

$$\kappa = \frac{P_o - P_e}{1 - P_e} \tag{1}$$

where P_o denotes the observed agreement and P_e represents the expected agreement by chance. Yielding values between -1 and 1, the Kappa coefficient provides a standardized measure of inter-rater reliability that adjusts for random agreement. Pruning redundant models with high mutual Kappa scores enhances ensemble variance while decreasing both spatial and temporal demands. Additionally, classifiers achieving below 75% accuracy on the test set were eliminated. This rigorous selection process, applied across 50 training iterations per architecture, ultimately yielded a refined subset of 25 diverse base models: 9 DenseNets, 7 ConvNexts, 5 EfficientNet-B0s, and 4 SEResNets. The pairwise Kappa scores corresponding to these selected models are illustrated in Fig. 9.

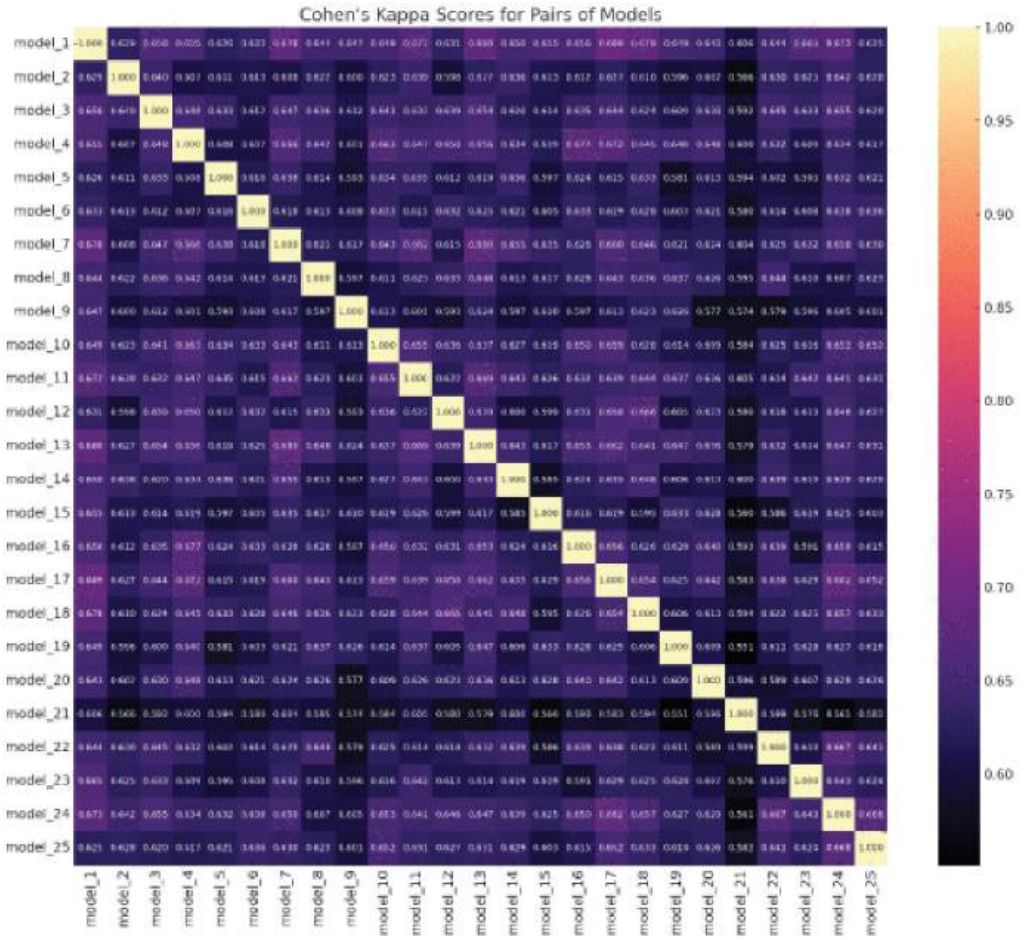


Figure 9. Pairwise Cohen's kappa coefficients calculated across the selected base models.

In the subsequent phase, the predictions generated by the retained base classifiers are synthesized. Within the proposed ensemble framework, a majority voting consensus is employed to integrate the outputs from the diverse array of base CNNs. Consequently, the final classification assigned by the ensemble is determined by the class label that receives the highest frequency of votes across the constituent models.

To optimize the foundational weights of the base models, the categorical cross-entropy (CCE) loss function was implemented, expressed as:

$$CCELoss = -\sum_{i}^{C} y_i \log(\hat{y}_i) \quad (2)$$

In this equation, $\hat{y}_i$ represents the model-predicted probability of a sample belonging to class i, while y_i serves as the binary ground-truth indicator. Given the seven distinct skin lesion categories in the dataset, the value of C is established as 7. To further refine the model and prevent overconfident predictions, label smoothing with a factor of $\alpha = 0.1$ was integrated as a regularization strategy. This technique modifies the traditional hard classification targets (0 and 1) by substituting them with values of $\frac{\alpha}{C-1}$ and $1 - \alpha$, respectively, thereby smoothing the softmax output distribution.

3.4 Training Setup

For the base model optimization, each network underwent training for 100 epochs utilizing the Adam optimizer with a learning rate of 0.001 and a batch size of 32; further

hyperparameters are detailed in Table 2, with remaining settings following PyTorch defaults. To ensure robustness, 50 independent training iterations were performed for each architecture, where each experiment utilized unique data subsets generated via the previously described weighted bootstrapping technique. Furthermore, network parameters were stochastically initialized in each run using the uniform Xavier initialization method [45].

Table 2. Experimental hyperparameters for base CNN training.

Optimizer	Batch size	Training Epochs	Learning rate
Adam	32	100	0.001

To mitigate the risk of overfitting, the validation loss was monitored throughout the training duration. Figure 10 illustrates representative plots of accuracy and loss metrics for both the training and validation cohorts. These specific visualizations correspond to two randomly selected instances from the 25 experiments that yielded the optimal models integrated into the final ensemble.

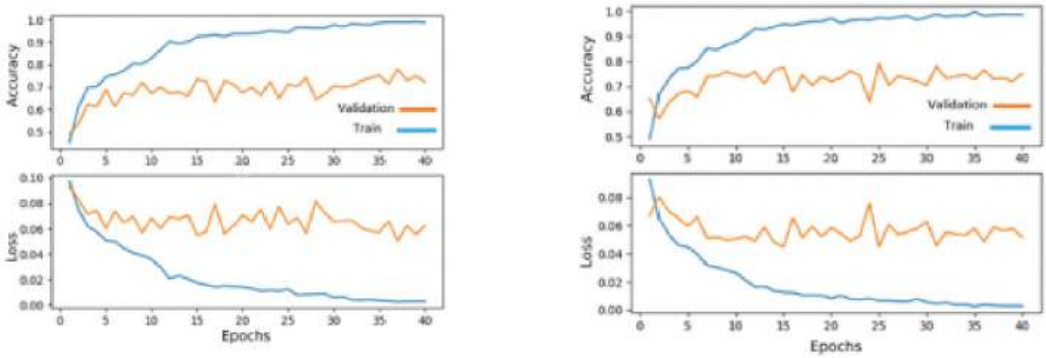


Figure 10. Training and validation curves for accuracy and loss.

4. | RESULTS

To assess the classification efficacy of the proposed framework, the model was trained using the ISIC 2018 training partition. Throughout the training phase, model performance was rigorously validated by monitoring the balanced accuracy on the 193-image ISIC 2018 validation cohort. The final performance metrics for the proposed ensemble classifier were then evaluated and reported using the independent ISIC 2018 test set. As demonstrated in Table 3, the proposed ensemble configuration consistently achieves superior accuracy compared to any of its individual constituent base classifiers.

Table 3. Mean accuracy and standard deviation for base models compared to the ensemble's overall performance.

Classifier	Mean Accuracy (%)	Standard Deviation
ConvNext-Tiny	77.6	5.2
EfficientNet-B0	74.59	6.3
SENet	73.79	4.54
DenseNet	73.58	3.9
ResNet50	70.32	4.2
Proposed Ensemble Accuracy (%)		90.15

Among the evaluated backbones, ConvNext-Tiny yielded the highest accuracy, demonstrating its efficacy in skin lesion classification, as detailed in Table 3. EfficientNet-B0 also achieved commendable mean accuracy; however, its higher standard deviation across 50

experiments indicates sensitivity to training data subsets, though its generalization remains robust. Conversely, SENet and DenseNet exhibited slightly lower mean accuracies but offered greater stability with lower standard deviations. Notably, SENet's balance between performance and consistency proves advantageous for this task, while ResNet produced the lowest mean accuracy despite maintaining a stable training process attributed to its residual connections. Ultimately, the proposed majority voting ensemble integrates 25 models from these four architectures, selected based on a 75% accuracy threshold and low mutual agreement to ensure diversity. The corresponding raw and normalized confusion matrices for the ensemble are illustrated in Fig. 11.

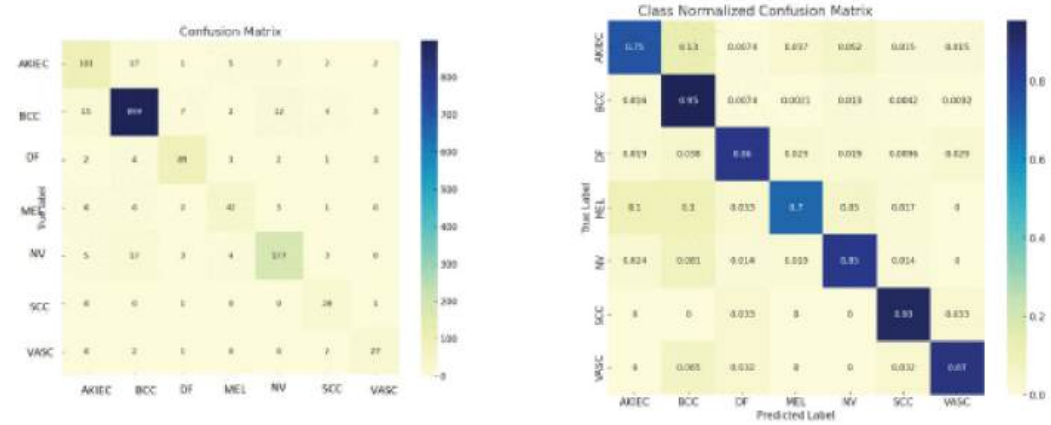


Figure 11. Confusion matrix (left) and class-normalized version (right) for the proposed ensemble on the test set, where diagonal elements indicate per-class accuracy.

Table 4 summarizes the total counts of true positives, false positives, true negatives, and false negatives. Furthermore, Fig. 12 and Table 5 present per-class performance metrics on the test set, specifically precision, recall, and F1-score.

Table 4. Aggregate counts of TP, FP, FN, and TN achieved by the ensemble model.

Type	TP	FP	FN	TN
Count	1363	149	149	1349

Table 5. Detailed per-class classification performance metrics on the ISIC 2018 test set for the proposed ensemble.

Class	Precision	Recall	F1-score
MEL	0.82	0.79	0.80
NV	0.96	0.97	0.96
BCC	0.85	0.85	0.85
AKIEC	0.80	0.73	0.76
BKL	0.78	0.79	0.78
DF	0.84	0.87	0.85
VASC	0.92	0.89	0.91

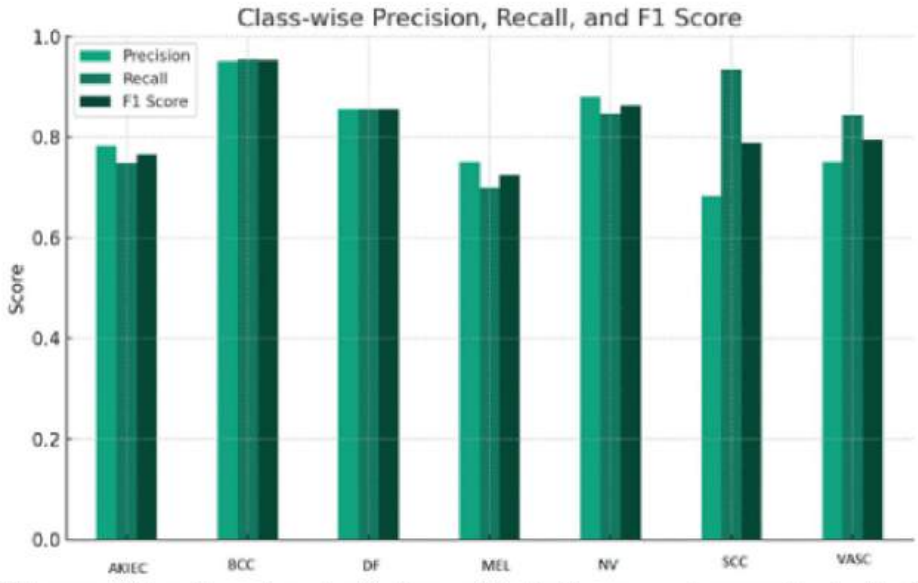


Figure 12. Representation of per-class classification metrics for the proposed approach using the ISIC 2018 test set.

A parametric evaluation of the base models is presented in Table 6, introducing an efficiency metric calculated as the ratio of accuracy to total parameter count (measured in millions). The results indicate that EfficientNet-B0 is the most optimized architecture, securing a competitive accuracy score despite maintaining the smallest parameter footprint among all evaluated base models.

Table 6. Comparison of test and validation accuracies alongside efficiency scores for the top-performing base models.

Base Model Name	Parameter Count (millions)	Mean Test Accuracy	Mean Validation Accuracy	Efficiency Score
ConvNext-Tiny	28.6	77.6	82.4	2.71
EfficientNet-B0	5.3	74.6	80.2	14.1
SENet	28.1	73.8	83.0	2.6
DenseNet121	8.0	73.6	79.1	9.2
ResNet50	25.6	70.3	77.7	2.75

As illustrated in Fig. 11, the proposed ensemble achieves its highest classification performance for Basal cell carcinoma, Nevus, and Dermatofibroma categories. Conversely, Melanoma represents the most challenging category, with approximately 20% of samples misidentified as Basal cell carcinoma or Actinic keratosis, indicating a predictive bias. Actinic keratosis follows as the second most difficult class, frequently misclassified as Basal cell carcinoma, Melanoma, or Nevus. Consequently, while the ensemble maintains high recall for most categories, it yields comparatively lower recall for the Basal cell carcinoma and Melanoma classes. Representative instances of these misclassified samples are presented in Fig. 13.

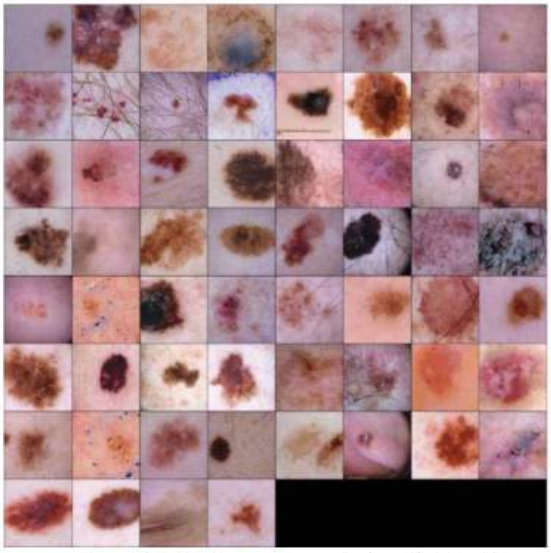


Figure 13. Representative examples of test set samples misclassified by the proposed ensemble model.

5. | ABLATION STUDY

To assess the individual contributions of the methodology's components, a sequence of experiments was conducted focusing on data augmentation, bootstrapping, kappa-score-based filtering of redundant models, and architectural diversity. The outcomes of these ablative studies are detailed in Table 7.

Table 7. Ablation study results evaluating the influence of data augmentation and bootstrapping on classification metrics.

Model	w/o bootstrapping and augmentation (Acc / F1)	Augmentation only (Acc / F1)	Augmentation and uniform bootstrapping (Acc / F1)	Augmentation and weighted bootstrapping (Acc / F1)	Filter by Kappa (Acc / F1)
DenseNet	68.32 / 65.0	68.1 / 65.7	70.3 / 67.9	73.6 / 69.2	73.6 / 69.2
EfficientNet B0	65.21 / 62.7	67.20 / 63.5	70.27 / 68.7	74.6 / 70.1	74.6 / 70.1
SENet	64.9 / 61.1	66.5 / 64.9	68.2 / 65.3	73.8 / 70.1	73.8 / 70.1
ConvNeXt Tiny	68.4 / 65.5	69.5 / 66.8	73.1 / 69.3	77.6 / 71.2	77.6 / 71.2
Ensemble	67.7 / 59.4	68.6 / 61.4	76.5 / 72.8	86.4 / 82.3	90.15 / 90.1

The ablation study highlights the substantial influence of data augmentation and bootstrapping on skin lesion classification performance. Baseline results without these techniques show moderate efficacy, with accuracy and F1-scores ranging from 64.9% to 68.4% and 61.1% to 65.5%, respectively. While augmentation alone marginally improves generalization, the most significant performance gains occur when it is paired with bootstrapping. Specifically, uniform bootstrapping enhances diversity among training subsets, while weighted bootstrapping further optimizes results by addressing class imbalance, providing a more equitable representation during training.

The most notable advancement is achieved through the ensemble approach, particularly when utilizing Kappa-score filtering to select the 25 least correlated base models. This strategy reduces redundancy among ensemble members and maximizes classifier-level diversity.

Consequently, the refined ensemble significantly outperforms both individual models and the initial configurations, reaching a peak accuracy of 90.15%. These findings underscore the efficacy of combining diverse, low-correlation models to enhance overall predictive robustness.

6. | DISCUSSION

Building upon existing skin lesion classification techniques [2], [15], [26], which primarily ensemble CNNs such as VGG and ResNet, this study trained multiple architectures on diversified, balanced, and augmented subsets of the ISIC 2018 dataset. Analysis identified ConvNext, SENet, DenseNet, and EfficientNet as the superior backbones. By integrating base classifiers with over 75% accuracy, a majority voting ensemble was established, achieving a robust 90.15% accuracy.

This voting strategy enhances predictive stability, while the class-balancing approach significantly improves the model's generalization across lesion types. Compared to competing methodologies [46], our streamlined deep learning framework remains highly competitive. As shown in Table 8, the proposed ensemble surpasses several fine-tuned ResNet and Inception-based models in balanced multi-class accuracy; however, while it outperforms many existing methods in overall accuracy, it remains 3% to 5% behind certain state-of-the-art approaches.

Table 8. Ablation study results evaluating the influence of data augmentation and bootstrapping on classification metrics.

Method	Uses external data	Balanced Multi-class accuracy	Accuracy
Skin Disease Classification by Resnet50 [48]	No	36.0%	83.9%
ResNet with triplet loss and feature similarity [49]	Yes	68.3%	95.3%
Ensembles of ResNet50, ResNet34, InceptionV3 [50]	No	72.8%	95.6%
Ensemble of VGG and GoogLeNet [51]	No	-	81.5%
Ensemble of different deep architectures and utilization of weighted loss [22]	No	-	85.1%
Conv. Ensemble (Inception Models Normalized+Un-Norm) + avg. pooling [18]	No	73.6%	93.6%
Maqsood and Damaševičius [52]	-	-	93.24%
Proposed Method	No	**84.0%**	90.15%

To strengthen the comparative analysis, the proposed methodology was further evaluated on the HAM10000 dataset employing a 5-fold cross-validation strategy. Results summarized in Table 9 indicate that the current approach exceeds existing methods in terms of accuracy, recall, and F1-score. Nevertheless, the precision achieved remains lower than that of the transformer-based architecture introduced by [47].

Table 9. Comparative performance analysis of various classification approaches on the HAM10000 dataset.

Author	Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Qian et al. [30]	DCNN-GMAB	91.6	-	73.5	-
Ali et al. [7]	EfficientNetB7	87.9	88	88	87
Srinivasu et al. [53]	DLNN + MobileNet V2 + LSTM	90.21	-	92.24	-
Shetty et al. [54]	ML + CNN	95.19	88	85	86
Wei et al. [55]	DenseNet + ConvNeXt	95.29	88.35	92.58	89.99
Nakai et al. [47]	Deep Bottleneck Transformer	95.84	**96.1**	-	-
Kousis et al. [56]	DenseNet169	92.25	-	-	93.27
Elashiri et al. [57]	S2C-DeLeNet	91.03	90.38	90.58	90.48
Datta et al. [31]	Inception ResNet v2 + Soft Attention	93.4	93.7	-	-
Tabibi et al. (Proposed Method)	Ensemble of ConvNext, DenseNet, SEResNet, EfficientNetB0	**96.12**	94.63	**93.11**	**93.86**

Despite the utilization of bootstrapping and data augmentation to address class imbalance and data diversity, sample scarcity persists for specific categories such as Nevus, Melanoma, and Actinic keratosis. Furthermore, the reliance on the ISIC 2018 dataset may introduce biases that limit global representativeness, necessitating the use of more diverse datasets in future research. Another promising direction involves the integration of vision transformer-based backbones, such as those introduced by [58], which are recognized for their ability to capture long-range dependencies and could further enhance ensemble performance.

Future investigations should also consider the strategic integration of Generative Adversarial Networks (GANs), as their efficacy in generating synthetic samples for minority classes is well-documented [59], [60], [61]. Optimizing GAN parameters alongside ensemble methods could significantly bolster dataset robustness. Additionally, refining data sampling techniques, exploring advanced augmentation, and incorporating patient demographic data may provide a more holistic and clinically applicable approach to skin lesion classification.

7. | CONCLUSION

This study establishes a foundation for future research by highlighting the critical role of architectural diversity and ensemble strategies in overcoming dataset complexities and class imbalances. The proposed ensemble, achieving a state-of-the-art accuracy of 90.15% on the test set, demonstrates that combining heterogeneous convolutional neural networks effectively captures a broad spectrum of features, significantly enhancing generalization. These findings suggest that the challenge of skin lesion classification is approaching a definitive solution, offering substantial promise for improving early dermatological diagnosis. Ultimately, the robustness of this framework confirms its clinical potential, with future advancements likely stemming from the integration of transformer-based classifiers into the ensemble.

REFERENCES

1. S. K. Datta, M. A. Shaikh, S. N. Srihari, and M. Gao, "Soft attention improves skin cancer classification performance," in Interpretability of Machine Intelligence in Medical Image Computing, and Topological Data Analysis and Its Applications for Medical Data, M. Reyes, P. Henriques Abreu, J. Cardoso, M. Hajij, G. Zamzmi, P. Rahul, and L. Thakur, Eds., Cham, Switzerland : Springer, 1007, pp. 13–23.
2. A. N. Omeroglu, H. M. A. Mohammed, E. A. Oral, and S. Aydin, "A novel soft attention-based multi-modal deep learning framework for multi-label skin lesion classification," Eng. Appl. Artif. Intell., vol. 120, Apr. 2023, Art. no. 105897.
3. P. Choudhary, J. Singhai, and J. S. Yadav, "Skin lesion detection based on deep neural networks," Chemometric Intell. Lab. Syst., vol. 230, Nov. 2022, Art. no. 104659.
4. G. Cirrincione, S. Cannata, G. Cicceri, F. Prinzi, T. Currieri, M. Lovino, C. Militello, E. Pasero, and S. Vitabile, "Transformer-based approach to melanoma detection," Sensors, vol. 23, no. 12, p. 5677, Jun. 2023.
5. L. I. Kuncheva, Combining Pattern Classifiers: Methods and Algorithms. Hoboken, NJ, USA : Wiley, 2014.
6. N. Codella, V. Rotemberg, P. Tschandl, M. Emre Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC)," 2019, arXiv:1902.03368.
7. ISIC Challenge2018, ISIC, 2018.
8. F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," Proc. IEEE, vol. 109, no. 1, pp. 43–76, Jan. 2021.
9. Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2022, pp. 11966–11976.
10. J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 7132–7141.
11. S. Mekruksavanich and A. Jitpattanakul, "Deep residual network for smartwatch-based user identification through complex hand movements," Sensors, vol. 22, no. 8, p. 3094, Apr. 2022.
12. M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. Int. Conf. Mach. Learn., 2019, pp. 6105–6114.
13. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 4700–4708.
14. PyTorch. (2024). DenseNet—PyTorch Hub. Accessed: Jul. 1, 2024. [Online]. Available: https://pytorch.org/hub/pytorch_vision_densenet/
15. X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in Proc. 13th Int. Conf. Artif. Intell. Statist., 2010, pp. 249–256.
16. ISIC. (2018). ISIC 2018 Challenge: Leaderboards. Accessed: Jan. 2024.
17. K. Nakai, Y.-W. Chen, and X.-H. Han, "Enhanced deep bottleneck transformer model for skin lesion classification," Biomed. Signal Process. Control, vol. 78, Sep. 2022, Art. no. 103997.

18. C. Mingang and W. Chen. (2018). Skin Disease Classification By Deep Network. [Online]. Available: https://isic-challenge-stade.s3.amazonaws.com/aa13d171-30f5-4622-b095-1a381f0d19d5/Skin_Disease_Classification_by_Deep_Network.pdf
19. Y. Xue, L. Gong, W. Peng, X. Huang, and Y. Zheng. (2018). Automatic Skin Lesion Analysis With Deep Networks. [Online]. Available: https://isic-challenge-stade.s3.amazonaws.com/4c9b36c3-161b-4b50-a820-92b85462f844/ISIC2018.pdf
20. A. H. Shahin, A. Kamal, and M. A. ElAttar. (2018). Automatic Skin Lesions Diagnosis With Deep Neural Networks. [Online]. Available: https://isic-challenge-stade.s3.amazonaws.com/1b91ec7e-1ab4-4440-ace0-4b30f219f8c7/Task3_article.pdf
21. T. Majtner, B. Bajić, S. Yildirim, J. Y. Hardeberg, J. Lindblad, and N. Sladoje, "Convolutional neural network based classification of dermoscopic images by transfer learning: ISIC 2018 challenge," ISIC Challenge Website Leaderboard Standing, Tech. Rep., 2022.
22. S. Maqsood and R. Damaševičius, "Multiclass skin lesion localization and classification using deep learning based features fusion and selection framework for smart healthcare," Neural Netw., vol. 160, pp. 238–258, Mar. 2023.
23. P. N. Srinivasu, J. G. Sivasai, M. F. Ijaz, A. K. Bhoi, W. Kim, and J. J. Kang, "Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM," Sensors, vol. 21, no. 8, p. 2852, Apr. 2021.
24. B. Shetty, R. Fernandes, A. P. Rodrigues, R. Chengoden, S. Bhattacharya, and K. Lakshmanna, "Skin lesion classification of dermoscopic images using machine learning and convolutional neural network," Sci. Rep., vol. 12, no. 1, p. 18134, Oct. 2022.
25. M. Wei, Q. Wu, H. Ji, J. Wang, T. Lyu, J. Liu, and L. Zhao, "A skin disease classification model based on DenseNet and ConvNeXt fusion," Electronics, vol. 12, no. 2, p. 438, Jan. 2023.
26. I. Kousis, I. Perikos, I. Hatzilygeroudis, and M. Virvou, "Deep learning methods for accurate skin cancer recognition and mobile application," Electronics, vol. 11, no. 9, p. 1294, Apr. 2022.
27. M. A. Elashiri, A. Rajesh, S. N. Pandey, S. K. Shukla, S. Urooj, and A. Lay-Ekuakille, "Ensemble of weighted deep concatenated features for the skin disease classification model using modified long short term memory," Biomed. Signal Process. Control, vol. 76, Jul. 2022, Art. no. 103729.
28. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16 × 16 words: Transformers for image recognition at scale," 2020, arXiv:2010.11929.
29. Z. Qin, Z. Liu, P. Zhu, and Y. Xue, "A GAN-based image synthesis method for skin lesion classification," Comput. Methods Programs Biomed., vol. 195, Oct. 2020, Art. no. 105568.
30. R. Kaur, H. GholamHosseini, and R. Sinha, "Synthetic images generation using conditional generative adversarial network for skin cancer classification," in Proc. IEEE Region 10 Conf. (TENCON), Dec. 2021, pp. 381–386.
31. K. W. Lee and R. K. Y. Chin, "The effectiveness of data augmentation for melanoma skin cancer prediction using convolutional neural networks," in Proc. IEEE 2nd Int. Conf. Artif. Intell. Eng. Technol. (IICAIET), Sep. 2020, pp. 1–6.
32. S. K. Datta, M. A. Shaikh, S. N. Srihari, and M. Gao, "Soft attention improves skin cancer classification performance," in Interpretability of Machine Intelligence in Medical Image

Computing, and Topological Data Analysis and Its Applications for Medical Data, M. Reyes, P. Henriques Abreu, J. Cardoso, M. Hajij, G. Zamzmi, P. Rahul, and L. Thakur, Eds., Cham, Switzerland : Springer, 1007, pp. 13–23.
33. A. N. Omeroglu, H. M. A. Mohammed, E. A. Oral, and S. Aydin, "A novel soft attention-based multi-modal deep learning framework for multi-label skin lesion classification," Eng. Appl. Artif. Intell., vol. 120, Apr. 2023, Art. no. 105897.
34. P. Choudhary, J. Singhai, and J. S. Yadav, "Skin lesion detection based on deep neural networks," Chemometric Intell. Lab. Syst., vol. 230, Nov. 2022, Art. no. 104659.
35. G. Cirrincione, S. Cannata, G. Cicceri, F. Prinzi, T. Currieri, M. Lovino, C. Militello, E. Pasero, and S. Vitabile, "Transformer-based approach to melanoma detection," Sensors, vol. 23, no. 12, p. 5677, Jun. 2023.
36. L. I. Kuncheva, Combining Pattern Classifiers: Methods and Algorithms. Hoboken, NJ, USA : Wiley, 2014.
37. N. Codella, V. Rotemberg, P. Tschandl, M. Emre Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC)," 2019, arXiv:1902.03368.
38. ISIC Challenge2018, ISIC, 2018.
39. F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," Proc. IEEE, vol. 109, no. 1, pp. 43–76, Jan. 2021.
40. Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2022, pp. 11966–11976.
41. J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 7132–7141.
42. S. Mekruksavanich and A. Jitpattanakul, "Deep residual network for smartwatch-based user identification through complex hand movements," Sensors, vol. 22, no. 8, p. 3094, Apr. 2022.
43. M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. Int. Conf. Mach. Learn., 2019, pp. 6105–6114.
44. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 4700–4708.
45. PyTorch. (2024). DenseNet—PyTorch Hub. Accessed: Jul. 1, 2024. [Online]. Available: https://pytorch.org/hub/pytorch_vision_densenet/
46. X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in Proc. 13th Int. Conf. Artif. Intell. Statist., 2010, pp. 249–256.
47. ISIC. (2018). ISIC 2018 Challenge: Leaderboards. Accessed: Jan. 2024.
48. K. Nakai, Y.-W. Chen, and X.-H. Han, "Enhanced deep bottleneck transformer model for skin lesion classification," Biomed. Signal Process. Control, vol. 78, Sep. 2022, Art. no. 103997.
49. C. Mingang and W. Chen. (2018). Skin Disease Classification By Deep Network. [Online]. Available: https://isic-challenge-stade.s3.amazonaws.com/aa13d171-30f5-4622-b095-1a381f0d19d5/Skin_Disease_Classification_by_Deep_Network.pdf

50. Y. Xue, L. Gong, W. Peng, X. Huang, and Y. Zheng. (2018). Automatic Skin Lesion Analysis With Deep Networks. [Online]. Available: https://isic-challenge-stade.s3.amazonaws.com/4c9b36c3-161b-4b50-a820-92b85462f844/ISIC2018.pdf
51. A. H. Shahin, A. Kamal, and M. A. ElAttar. (2018). Automatic Skin Lesions Diagnosis With Deep Neural Networks. [Online]. Available: https://isic-challenge-stade.s3.amazonaws.com/1b91ec7e-1ab4-4440-ace0-4b30f219f8c7/Task3_article.pdf
52. T. Majtner, B. Bajić, S. Yildirim, J. Y. Hardeberg, J. Lindblad, and N. Sladoje, "Convolutional neural network based classification of dermoscopic images by transfer learning: ISIC 2018 challenge," ISIC Challenge Website Leaderboard Standing, Tech. Rep., 2022.
53. S. Maqsood and R. Damaševičius, "Multiclass skin lesion localization and classification using deep learning based features fusion and selection framework for smart healthcare," Neural Netw., vol. 160, pp. 238–258, Mar. 2023.
54. P. N. Srinivasu, J. G. Sivasai, M. F. Ijaz, A. K. Bhoi, W. Kim, and J. J. Kang, "Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM," Sensors, vol. 21, no. 8, p. 2852, Apr. 2021.
55. B. Shetty, R. Fernandes, A. P. Rodrigues, R. Chengoden, S. Bhattacharya, and K. Lakshmanna, "Skin lesion classification of dermoscopic images using machine learning and convolutional neural network," Sci. Rep., vol. 12, no. 1, p. 18134, Oct. 2022.
56. M. Wei, Q. Wu, H. Ji, J. Wang, T. Lyu, J. Liu, and L. Zhao, "A skin disease classification model based on DenseNet and ConvNeXt fusion," Electronics, vol. 12, no. 2, p. 438, Jan. 2023.
57. I. Kousis, I. Perikos, I. Hatzilygeroudis, and M. Virvou, "Deep learning methods for accurate skin cancer recognition and mobile application," Electronics, vol. 11, no. 9, p. 1294, Apr. 2022.
58. M. A. Elashiri, A. Rajesh, S. N. Pandey, S. K. Shukla, S. Urooj, and A. Lay-Ekuakille, "Ensemble of weighted deep concatenated features for the skin disease classification model using modified long short term memory," Biomed. Signal Process. Control, vol. 76, Jul. 2022, Art. no. 103729.
59. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16 × 16 words: Transformers for image recognition at scale," 2020, arXiv:2010.11929.
60. Z. Qin, Z. Liu, P. Zhu, and Y. Xue, "A GAN-based image synthesis method for skin lesion classification," Comput. Methods Programs Biomed., vol. 195, Oct. 2020, Art. no. 105568.
61. R. Kaur, H. GholamHosseini, and R. Sinha, "Synthetic images generation using conditional generative adversarial network for skin cancer classification," in Proc. IEEE Region 10 Conf. (TENCON), Dec. 2021, pp. 381–386.
62. K. W. Lee and R. K. Y. Chin, "The effectiveness of data augmentation for melanoma skin cancer prediction using convolutional neural networks," in Proc. IEEE 2nd Int. Conf. Artif. Intell. Eng. Technol. (IICAIET), Sep. 2020, pp. 1–6.