Journal of Information Hiding and Multimedia Signal Processing ©2025 ISSN 2073-4212
Ubiquitous International Volume **16**, Number **1**, March **2025**

A Classification Method based on Concatenation Features for Diagnosing Skin Diseases

Long Duong Phi[1,2]*, Binh Pham Nguyen Thanh[1,2], and Quang Tran Van[1,2]

[1] University of Information Technology, Vietnam
[2] Vietnam National University Ho Chi Minh City, Vietnam
{longdp, binhpnt}@uit.edu.vn, 20520722@gm.uit.edu.vn

* Corresponding author: Long Duong Phi
Received November 18, 2024, revised January 17, 2025, accepted January 20, 2025.

Abstract. *Negative changes in weather and living environments have led to a global increase in dermatological cases. Skin diseases and skin cancer significantly impact human health, mental well-being, and life. Most skin disease and skin cancer cases can be completely treated if detected early, highlighting the necessity for early screening and diagnosis. In recent years, deep learning (DL)-based dermatological image classification methods have achieved remarkable progress. In this study, convolutional neural networks (CNNs) are used to classify seven groups of skin diseases and cancers in the ISIC2018 dermoscopy image dataset. Geometric data augmentation is applied to mitigate class imbalance within the dataset. Transfer learning methods, including ResNet-50, VGG-19, and EfficientNet-V2, are experimentalized to evaluate classification quality. Additionally, the study combines these three models using a Concatenation Features method (CF) to propose a new classifier. Experimental results indicate that the new CF model outperforms previous models, achieving an accuracy of 89.80%, an F1-score of 92.23%, and stable classification quality across all disease classes.*
Keywords: Skin Disease; Image Classification; Deep Learning; Ensemble Method; Concatenate Features.

1. **Introduction.** Skin diseases and skin cancers are changes in the skin, usually due to abnormal proliferation of epidermal cells. In recent years, the incidence of skin cancer has increased significantly worldwide, with millions of cases of non-melanoma skin cancer and hundreds of thousands of cases of melanoma skin cancer reported each year. Ozone depletion and climate change have increased the risk of exposure to ultraviolet radiation, leading to a higher risk of skin diseases, especially cases related to benign tumors and melanin pigment lesions. Therefore, identifying and classifying skin diseases and skin cancers becomes important, supporting timely diagnosis and treatment [1].

Early detection of skin diseases and skin cancers, especially melanoma, is very important to improve patient survival rates and reduce treatment costs [2]. Late-stage melanoma has a high mortality rate, with more than 20,000 people dying from melanoma in Europe each year, however, if detected at an early stage, melanoma and skin diseases can have better treatment regimens and higher recovery rates. In addition, skin diseases and skin cancers have a major impact on quality of life, both physically and mentally, due to their visible nature and chronic symptoms.

However, the diagnosis of skin diseases and cancers faces a major challenge due to the diversity in the manifestations of skin lesions, including changes in color, surface texture,

and shape [3]. Traditional diagnostic methods rely heavily on the expertise and experience of dermatologists, leading to potential human errors and inconsistencies in disease diagnosis. With the rapid development of artificial intelligence and deep learning technologies, computer-aided diagnosis (CAD) systems are proving useful in medical image analysis [4]. These computer systems can process large amounts of data and identify patterns that may not be immediately visible to the human eye. Deep learning, especially convolutional neural networks (CNNs), has demonstrated superior performance in image classification tasks, making it an ideal solution for the complex and nuanced field of dermatology. However, the current single CNN networks use a single convolutional neural network, which results in a rather low classification performance, because each CNN network design has its own advantages and disadvantages on each separate data layer, so the ability to extract features on large datasets in the medical field of a single CNN network is not guaranteed in most cases.

In this study, we propose a method that concatenate features extracted from deep learning networks, specifically convolutional neural networks (CNNs), to enhance the accuracy and efficiency of diagnosing skin diseases and skin cancer. We utilize high-quality dermoscopy images from the ISIC2018 dataset and apply data augmentation techniques to tackle the challenges posed by imbalanced datasets, thereby improving classification performance. Our ultimate goal is to develop a reliable diagnostic classifier that can assist dermatologists in making quicker and more accurate decisions.

2. **Related work.** The initial detection of traditional skin diseases was evaluated visually by humans, primarily using the pattern analysis method proposed by Pehamberger et al. [5], the ABCD rule proposed by Stolz [6], and the seven-point detection method proposed by Argenziano [7]. These methods fundamentally rely on manual analysis of specific characteristics to support diagnosis. While they can achieve good diagnostic accuracy, they are energy-intensive and have their limitations. Dermatoscopy, with its ability to provide detailed images, helps improve the accuracy of diagnosis. However, dermoscopic image-based diagnosis remains challenging due to the diversity and complexity of skin lesions. To overcome this problem, automated computer-aided diagnosis (CAD) systems for skin diseases and skin cancers are increasingly being used [4].

Today, with the advancement of computer technology, computer-aided diagnosis has been introduced in the field of image detection through dermoscopy. This diagnostic system can screen a large number of dermoscopic images and provide diagnostic opinions within an acceptable error range, helping to reduce the workload of dermatologists and minimize unnecessary pathological analysis [4]. In 2012, Rahil Garnavi et al. presented a novel computer-aided diagnosis system for skin diseases and skin cancer by optimally selecting and integrating features derived from the texture, contour, and geometric characteristics of lesions and classifying them using Support Vector Machine, Random Forest, Logistic Model Tree, and Hidden Naive Bayes methods [8]. Through image data preprocessing to remove noise and unwanted hair, followed by applying Support Vector Machine and K-Nearest Neighbors algorithms for feature extraction, in 2015, R. Sumithra et al. proposed a new method for the automatic segmentation and classification of skin lesions [9]. Utilizing texture, color, and geometric features, Hatem proposed a system that can identify and classify skin lesions and skin cancer based on the K-Nearest Neighbors algorithm in 2022 [10].

Classifying skin diseases and skin cancer using traditional machine learning methods can achieve certain results, but there are still limitations that prevent them from fully ensuring accuracy. The disadvantages include poor performance with large datasets and sensitivity to noise. Additionally, images in the trained dataset may face difficulties in

identifying appropriate features, and achieving high effectiveness remains a significant challenge. Furthermore, machine learning methods for classifying skin diseases and skin cancer need to be trained on training data, but their effectiveness and accuracy depend on new, untrained data. Therefore, in the near future, these methods are gradually being replaced by deep learning methods, and results have shown that deep learning offers improvements in performance and accuracy [4]. Convolutional Neural Networks (CNNs) have been applied and achieved much success in computer vision tasks and have been used for diagnosing skin diseases. CNNs can recognize complex features from images and create feature maps, thereby assisting dermatological experts in quickly diagnosing skin diseases with high accuracy. In 2017, Andre Esteva and colleagues applied CNNs to classify skin cancer, using advanced methods like ResNet and Inception to train the models, with results showing high performance in classifying common skin diseases and skin cancers [11]. By combining advanced methods such as ResNet-50 and DenseNet-12 with Neural Architecture Search (NAS) techniques to optimize model performance, in 2018, Jainesh Rathod and his team announced an automated skin disease diagnosis system based on CNNs, achieving high accuracy. Generative Adversarial Networks (GANs) have also been used to generate synthetic images to augment training data for deep learning models, improving model accuracy and generalization [12]. In 2019, Xin Yi and colleagues provided a new perspective on high-performance methods, especially in the context of a lack of labeled data, opening new directions for applying deep learning in skin disease diagnosis [13]. In 2023, Gomathi et al. proposed a novel dual optimization-based deep learning network (DODL net) for skin cancer detection. The DODL net employs a hybrid approach, combining Bacterial Foraging Optimization (BFO) and Particle Swarm Optimization (PSO) algorithms to extract features from segmented images [14]. In 2023, Karpagam Mridha et al. proposed a method that employs a Gaussian filter to remove noise from skin images and utilizes Support Vector Machine for differentiating and detecting nevus and melanoma skin diseases [15].

Through previous studies, it can be seen that the ensemble methods have better performance and classification accuracy than the single research models, based on that advantage, many research groups have tried to take advantage of ensemble models to improve the classification quality. Based on the ISIC 2017 image dataset of skin lesions, in 2020, Nils Gessert et al. also proposed a classification model based on the synthesis of three models EffientNets, SENet and ResNeXt WSL [16]. In 2022, Mohamed A. Elashiri and his colleagues proposed a weighted deep feature concatenating fusion method with Modified Long Short Term Memory (MLSTM) based on the training results of three individual CNN models DeepLab-v3, ResNet-50, and VGG-16, and the weights were optimized using proposed Hybrid Squirrel Butterfly Search Optimization (HSBSO) [17]. In 2023, Minarul Islam Raju and colleagues proposed the RFSVCLR method, an ensemble machine learning algorithm by stacking Random Forest, Support Vector Machine, and Logistic Regression for chronic kidney disease prediction. This algorithm shows better results than other classifiers [18]. In 2024, Abdallah Radwan et al. introduced an ensemble classification for coronary artery disease prediction [19]. They used SMOTE-ENN and applied various data pre-processing techniques to solve the imbalance in the dataset. Afterward, they experimented with and evaluated eight individual classification models and multi-classifier fusion. Their experimental results show that ensemble classification between Logistic Regression, Light Gradient Boosting (LightGBM), and Adaptive Boosting (AdaBoost) is better than the individual classifier. These research and experimental findings have paved the way for a novel approach to improving classification performance by fusing individual classifiers.

3. **Methodology.** In the problem of skin disease image classification, most previous studies have trained on a single CNN model, which often has its outstanding features on a separate data group, so the results obtained are often uneven on the data labels. In this study, we propose a classification model that improves feature extraction and increases classification performance by combining three base CNN classifiers, including ResNet-50, VGG-19, and EfficientNet-V2. These CNNs were chosen for their advantages and proven high performance in various image classification tasks, including complex medical datasets like MRIs, X-rays, and dermoscopic images. These models have been optimized and pre-trained on the ImageNet dataset, enhancing their generalization capabilities for real-world medical applications. The classifiers have been trained, evaluated, and integrated into a Concatenation Features (CF) method. The overall architecture of the proposed approach is illustrated in Figure 1.

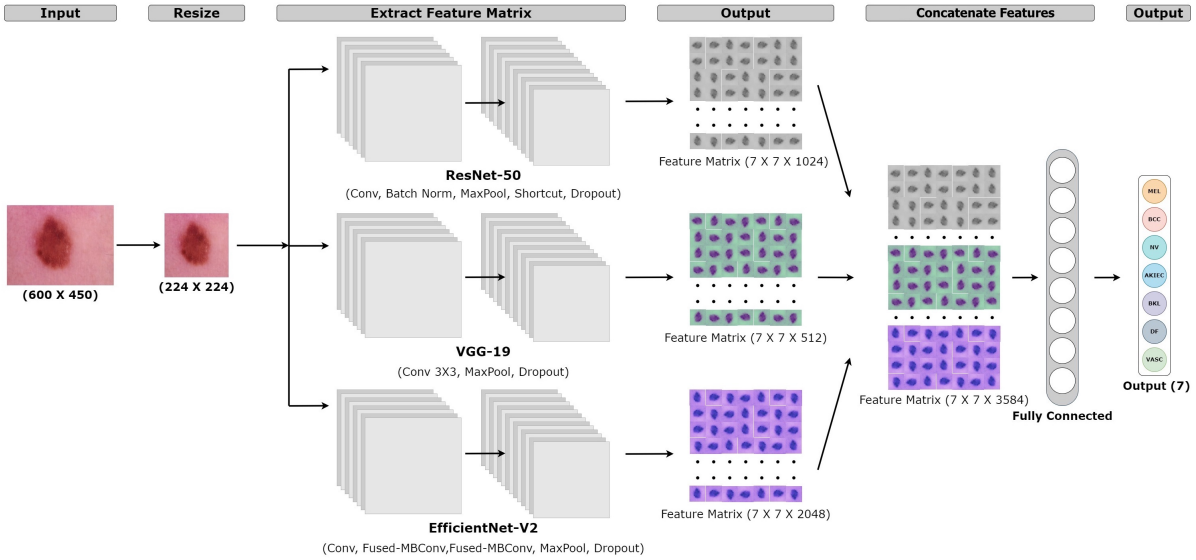

FIGURE 1. The overall architecture of the Concatenation Features for Skin Disease Diagnosis

3.1. **Base CNN classifier.**

3.1.1. *ResNet-50.* Residual Network (ResNet) introduced by Kaiming He et al. at Microsoft Research, is a breakthrough in deep learning, helping to improve the accuracy in image recognition through the residual learning mechanism [20]. ResNet belongs to the CNNs family and is designed to work with hundreds of layers. However, as the number of convolutional layers increases, the vanishing gradient phenomenon makes the learning process inefficient. The authors have overcome this by introducing residual blocks, which use shortcut connections to maintain stable gradients, improving the learning ability and performance of the network. Variants of ResNet, such as ResNet-18, ResNet-34, ResNet-50, and ResNet-101, are distinguished by the number of layers. ResNet-50, with 50 layers and a hierarchical structure containing many residual blocks, has demonstrated superior performance on the ImageNet dataset.

3.1.2. *VGG-19.* The VGG network, developed by the Visual Geometry Group at the University of Oxford, has produced several successful methods, such as VGG-16 and VGG19, with practical applications in tasks such as face recognition and image classification [21].

The main goal of VGG research is to explore how the depth of the network affects classification accuracy on large-scale data. Unlike previous CNNs that used large kernels, which resulted in many parameters and training challenges, VGG-19 uses small (3×3) convolutional layers stacked on top of each other to increase depth. The model has 19 layers, including 16 convolutional layers and 03 fully connected layers, with pooling layers to reduce the dimensionality of the feature. When tested on the ImageNet dataset, VGG-19 achieved high accuracy, surpassing many previous models.

3.1.3. *EfficientNet-V2.* EfficientNet, introduced by Mingxing Tan and Quoc V. Le, is a high-performance CNNs model that incorporates a compound scaling method to balance accuracy and computational resources [22]. It stands out for its reduced number of parameters and FLOPS and provides an architectural blueprint using neural architecture search to design a new baseline network and scale it up to obtain a family of models called EfficientNets, which achieve much better accuracy and efficiency than previous CNNs. EfficientNet-V2 also introduced by Mingxing Tan and Quoc V. Le, which speeds up training while maintaining accuracy. EfficientNet-V2 improves by combining new MB-Conv and Fused-MBConv, using a smaller scaling ratio and 3×3 kernel layers. These changes reduce memory access costs and optimize training performance [23].

3.2. **Concatenation Features (CF).** During the experiment of CNN models on the dermoscopy image dataset, our research team found that the best classification results for each disease class were various. The classification performance varies according to the characteristics of each class, so combining multiple models can improve the quality and performance of the classification. The basic CNN network consists of 3 layers: Convolutional, Pooling, and Fully Connected. CNN methods all use multiple Convolutional layers to extract features from the image. The final feature map, containing high-level features, makes the combined model more comprehensive.

Suppose, a CNN deep learning model has the output of the convolutional layer as a feature map F_i with dimensions (H_x, W_y, C_z) where H_x is the height, W_y is the width, and C_z is the number of channels or depth of the feature map. To combine feature maps together, methods such as concatenation, stacking, or pooling can be used. In this study, we use the method of concatenating feature maps to create a new feature map. The concatenation method is chosen for feature fusion because it can fully preserve the information extracted from each individual model. This approach contrasts with other fusion techniques, such as addition or averaging, which may risk losing critical information during integration.

$$F_C = Concatenate(F_1, F_2, ..., F_i) \tag{1}$$

Where:
- F_C**:** is the newly concatenated feature map.
- $F_1, F_2, ..., F_i$**:** are the feature maps used for concatenation.

Through the research and experimentation, we observed that the three models, ResNet-50, VGG-19, and EfficientNet-V2, complement each other effectively. Specifically, ResNet-50 utilizes residual blocks to capture deeper features and solve the vanishing gradients problem. VGG-19, with its sequential structure and greater depth, extracts low-level features, focusing on local details. EfficientNet-V2 integrates an efficient scaling mechanism (compound scaling) to extract more complex features while maintaining high computational efficiency. The CF model is formed by combining the outputs of three

trained models: VGG-19, ResNet-50, and EfficientNet-V2. The input is a tensor of size (1, 3, 224, 224) representing the RGB image, which is processed in evaluation mode. Each model produces outputs, denoted $\mathbf{y}_A$, $\mathbf{y}_B$, and $\mathbf{y}_C$, with different sizes corresponding to $\mathbb{R}^{m_A}$, $\mathbb{R}^{m_B}$, $\mathbb{R}^{m_C}$. These outputs are then flattened and concatenated into a composite output vector:

$$\mathbf{y_{concatenate}} = [\mathbf{y}_A; \mathbf{y}_B; \mathbf{y}_C] \in \mathbb{R}^{m_A+m_B+m_C} \tag{2}$$

In this study, the number of skin disease classes to be classified is 7. The training of the previous individual models yields feature matrices of size VGG-19: $\mathbf{y}_A = [7 \times 7 \times 512]$, ResNet-50: $\mathbf{y}_B = [7 \times 7 \times 1024]$ and EfficientNet-V2: $\mathbf{y}_C = [7 \times 7 \times 2048]$ respectively, the feature map are concatenated via formula (2) to obtain a new matrix of size $[7 \times 7 \times 3584]$, depth concatenation matrix. Then, the newly concatenation feature map will pass through the Fully Connected layer and produce a classification result as one of the disease labels. This vector is fed into a classifier layer with the weight matrix W and bias b, n is the number of classes to be classified, producing a final output of size:

$$\mathbf{output} = \mathbf{W} \cdot \mathbf{y_{concatenate}} + \mathbf{b} \in \mathbb{R}^{1 \times n} \tag{3}$$

In conclusion, we propose the CF model by concatenating the feature maps after being trained by three individual models. First, train the models separately with the augmented dermoscopy image dataset, then model will be retrained, where the weights of the three individual models are frozen during training, and only the weight of the classification layer is optimized. This process allows the CF model to take advantage of different features from the three individual models to improve the accuracy of the classification task.

4. Experimental Evaluation.

4.1. Dataset and data pre-processing.

This study used the ISIC2018 dataset from The International Skin Imaging Collaboration (ISIC) [24], which provides a large collection of 11,720 dermoscopy images labeled according to 7 skin disease types, including Actinic Keratoses (AKIEC), Basal Cell Carcinoma (BCC), Dermatofibroma (DF), Melanoma (MEL), Benign Keratosis (BKL), Melanocytic nevi (NV), Vascular Skin Lesions (VASC). This structured and widely used dataset is a standard reference for research on dermatological image analysis. The dataset contains two file types:

- CSV: This file contains image identification data and disease labels across 7 classes (AKIEC, BCC, DF, MEL, BKL, NV, VASC). Each image belongs to one class, represented by a "1" or a "0".

- JPG: The images, 600 × 450 pixels, were captured using dermoscopy and labeled according to common lesion types.

The images were resized from 600 × 450 to 224 × 224 pixels to fit the CNN models. Due to data imbalance, where NV accounts for 66% of the data and DF and VASC account for less than 2%, data augmentation techniques (rotation, translation, scaling, flipping, cropping) were applied to AKIEC, BCC, MEL, VASC labels on the training and validation sets. The dataset was split into training, validation, and testing sets in a ratio of 8:1:1, dataset after augmentation as shown in **Table 1**. Additionally, the mean and standard deviation values determined for each color channel were also used to normalize the dataset, as shown in **Table 2**.

4.2. Results and Discussion.

In this study, we trained and evaluated the classification performance of classifiers on the ISIC2018 dataset. We also tuned the hyper-parameters to suit the size of the dataset and the hardware configuration to achieve the optimal

Table 1. Number of images per class in each train, validation, and test sets

Class	Train set		Validation set		Test set
	Original	Augmentation	Original	Augmentation	
AKIEC	302	1,208	37	148	39
BCC	497	994	62	124	63
DF	1,070	1,070	133	133	135
MEL	128	896	16	112	16
BKL	1,044	1,044	130	130	131
NV	6,189	6,189	773	773	775
VASC	144	1,008	18	126	18
Total	9,374	12,409	1,169	1,546	1,177

Table 2. Data pre-processing parameters for the individual classifiers

	ResNet-50	VGG-19	Efficient-V2	CF
Input size	[224 × 224]			
mean	[0.485, 0.456, 0.406]			
std	[0.229, 0.224, 0.225]			

results, as shown in **Table 3**. The experimental platform is Kaggle, using 2 NVIDIA T4 GPUs, 30GB RAM.

Table 3. Hyper-parameters of the individual classifiers

	ResNet-50	VGG-19	Efficient-V2	CF
Num-epochs	64	64	64	32
Batch size	32			
Optimizer	Adam			
Loss function	Cross-Entropy Loss			
Activation function	Sigmoid	Sigmoid	SiLU	Sigmoid
Learning rate	weight_decay= 0.0002, lr=0.0001			

In the experiment, three individual models VGG-19, ResNet-50, and EfficientNet-V2 are loaded from the PyTorch library with pre-trained weights on ImageNet dataset (version IMAGENET1K-V1) as initializers. The models are adapted for the 7-class classification task by replacing the last classification layer. Specifically, the original classification layer of each model is replaced by a sequence of new layers, including two Dropout layers with probability 0.5, a linear layer reducing the output dimensionality to 512, a nonlinear activation function (*Sigmoid* for VGG-19 and ResNet-50, *SiLU* for EfficientNet-V2), and a final linear layer for 7-class classification. All three models use the *Adam* optimization algorithm with a fixed learning rate of 0.0001, betas of (0.9, 0.999), and weight decay of 0.0002 for VGG-19 and ResNet-50, and 0.0001 for EfficientNet-V2, to reduce overfitting. For the CF method, the model will be retrained with the weights of the individual models frozen, and only the weights in the classification layer will be optimized through the *Adam* algorithm with a learning rate of 0.0001 and weight decay of 0.0002. To make the most of the hardware, the models apply model parallelization (*torch.nn.DataParallel*) on multiple GPUs if available. Finally, the models are brought to the computing platform (*device*) to optimize performance during training and inference.

The models were evaluated using five commonly used performance metrics: Accuracy, Weighted Average Precision (WAvg Precision), Weighted Average Recall (WAvg Recall), and Weighted Average F1-score (WAvg F1-score). These metrics offer a holistic view of each model's ability to classify instances, assess their overall effectiveness, and provide important insights into the models' strengths and limitations.

Table 4. Comparison of Accuracy, WAvg Precision, WAvg Recall, and WAvg F1-score for 3 individual models and Concatenation Features (CF) model

Model	**Accuracy**	**WAvg Precision**	**WAvg Recall**	**WAvg F1-score**
ResNet-50	83.01%	82.80%	83.01%	82.77%
VGG-19	86.07%	85.96%	86.07%	85.85%
EfficientNet-V2	87.85%	87.44%	87.85%	87.55%
CF	**89.80%**	**92.47%**	**91.18%**	**92.23%**

The results of the experiments on the deep learning methods on the skin disease and skin cancer image dataset using data augmentation techniques are shown in **Table 4**. The results of the WAvg Recall, WAvg Precision, Accuracy, and WAvg F1-score indices show that among the three individual deep learning models, EfficientNet-V2 is the model with the highest overall performance with Accuracy of 87.85% and the highest WAvg Recall, WAvg Precision and WAvg F1-score. Also, this experiment found that the new CF model developed by the research team performs significantly better than the individual models. The CF model has an Accuracy of 89.80%, an increase of 2.05% compared to the highest method among the three single models, EfficientNet-V2, the WAvg F1-score also increased significantly with 92.23%, 4.68% higher than EfficientNet-V2, the WAvg Precision reached 92.47% and finally WAvg Recall reached 91.18%, which also significantly improved compared to 3 individual models.

The test results on each disease label are shown in **Table 5**, showing the difference in classification for each type of skin disease and skin cancer of individual models. When evaluating the performance of models, it is evident that all three individual models perform well on the NV class, which is the majority class in the dataset, with Precision, Recall, and F1-score all above 90%. ResNet-50 achieves stable performance with a difference between Precision and Recall across disease classes. For the AKIEC, MEL, and VASC classes, the model shows good Precision, but Recall is lower, indicating that the model may miss some cases from these classes, which impacts its overall performance. However, ResNet-50 demonstrates good detection for the BCC, DF, and BKL classes, especially with DF, where Recall is quite high. VGG-19 performs reasonably well in the BCC and VASC classes but performs poorly in AKIEC, MEL, and BKL classes, where both Precision and Recall are low, particularly for AKIEC and MEL. While VGG-19 has high Precision for the DF class, its Recall is low. These results suggest that VGG-19 misses many cases and has not yet achieved a good balance between Precision and Recall. EfficientNet-V2 shows the highest performance among the three individual models, achieving the highest F1-score across all 7 disease classes, particularly for classes like NV and VASC, where the model achieves very high Precision and Recall, resulting in impressive F1-scores. Although EfficientNet-V2 demonstrates a better balance between Precision and Recall for the AKIEC, BCC, and VASC classes, it still struggles with others. For instance, the model has high Precision but low Recall on the DF class, leading to a less-than-optimal F1-score. Despite correctly classifying most DF cases, the model misses a few and needs improvement in detection ability.

The test results of the CF model show that the classification performance on each disease label has an improvement in the metric difference, specifically comparing the F1-score of the CF model with the EfficientNet-V2 model, which is the model with the highest F1-score of the three individual models. The F1-score of the CF model is higher than EfficientNet-V2 by 1.3%, 4.07%, 5.8%, 7.88%, 8.1%, 16.63% across the NV, AKIEC, MEL, BCC, DF, and BKL classes. VASC is the only class with a slight performance drop, but compared to ResNet-50 or VGG-19, it still represents a significant performance improvement. For the majority of disease classes, the CF model outperforms the individual models in terms of Precision and Recall. The CF model's precision for the VASC class is less than that of three individual models, but its Recall is excellent, equal to EfficientNet-V2, and outperforming the two others. Although the CF model has achieved a better balance between Precision and Recall across most disease classes, the disparity remains large in the MEL class. This indicates that the CF model still faces some challenges in accurately classifying this particular class.

During the experiment, we monitored and recorded the training time of the models and the average time to diagnose an image in the test set, the recorded results are shown in **Table 6**. Among the individual models, it can be seen that EfficientNet-V2 has the longest training time of 282 minutes and 19 seconds and is also the model with the highest performance among the individual models. The result showing the fastest learning and convergence speed is ResNet-50 with 111 minutes 44 seconds, also giving positive diagnosis results. As presented earlier, the CF model freezes the weights of the individual models and continues training to optimize the weight of the classification layer. The additional training time required is 152 minutes and 58 seconds. Regarding testing performance, all models exhibit extremely quick prediction times. The EfficientNet-V2 model and ResNet-50 both predict results quickly, taking about 0.0111 seconds per image. Compared to ResNet-50 and EfficientNet-V2, VGG-19 has a slightly longer prediction time of 0.0139 seconds per image. The CF model has a longer prediction time of 0.0197 seconds per image, which may be due to the combination of individual models requiring additional computation during the prediction phase.

In conclusion, the CF model significantly improved the classification performance. Specifically, 4 out of 7 classes achieved a performance increase of more than 5%, and one class achieved an increase of more than 10% when compared to the best results from the individual models. Although there was a slight decrease in one class (1.3%), the CF model still outperformed. Experimental results showed that this concatenation features approach outperformed the individual models, as demonstrated by F1-scores metric in many different classes. Although models like EfficientNet-V2 and ResNet-50 also performed well, the CF model demonstrated an exceptional combination of detecting true positive cases (Recall) and achieving accurate classification (Precision) across most disease classes. This indicates that the CF model has a high classification performance, helping to minimize false positives, especially in classes prone to confusion. Besides that, CF outperformed at detecting true positive cases, minimizing false negatives, both in more common classes (such as NV) and less common ones (such as BCC and BKL). The reason for this improvement is that the CF model combines the strengths and a diverse and rich set of features from the three individual models, creating a set of feature maps with greater detail and richness. Therefore, the CF model is the most outstanding and well-balanced model when considering both Precision and Recall, making it a powerful and efficient model for classifying and detecting skin diseases in dermatological image data. Although the training time for the CF model is longer, the prediction time for a single image is not significantly different from that of the individual models. In medical image classification tasks, the balance between accuracy and execution time depends on

TABLE 5. Comparison of Confusion matrix and Precision, Recall, and F1-score on disease labels between 3 individual models and CF model

Class	**ResNet-50**									
	AKIEC	BCC	DF	MEL	BKL	NV	VASC	**Precision**	**Recall**	**F1-score**
AKIEC	24	7	1	2	6	0	0	80.00%	61.54%	69.57%
BCC	3	49	1	1	6	3	0	73.13%	77.78%	75.38%
DF	0	1	12	1	0	2	0	66.67%	75.00%	70.59%
MEL	2	1	1	77	16	34	0	72.64%	58.78%	64.98%
BKL	1	2	1	7	106	18	0	72.20%	78.52%	74.13%
NV	0	6	3	18	17	730	1	92.52%	94.19%	93.35%
VASC	0	1	0	0	0	2	15	93.75%	83.33%	88.24%
Class	**VGG-19**									
	AKIEC	BCC	DF	MEL	BKL	NV	VASC	**Precision**	**Recall**	**F1-score**
AKIEC	19	2	0	4	12	2	0	59.38%	48.72%	53.52%
BCC	5	48	1	2	3	4	0	81.36%	76.19%	78.69%
DF	2	1	8	0	0	5	0	88.89%	50.00%	64.00%
MEL	2	1	0	78	13	37	0	56.52%	59.54%	57.99%
BKL	3	3	0	19	86	24	0	68.80%	63.70%	66.15%
NV	0	4	0	35	11	724	1	90.61%	93.42%	91.99%
VASC	1	0	0	0	0	3	14	93.33%	77.78%	84.85%
Class	**EfficientNet-V2**									
	AKIEC	BCC	DF	MEL	BKL	NV	VASC	**Precision**	**Recall**	**F1-score**
AKIEC	31	2	0	2	2	2	0	79.49%	79.49%	79.49%
BCC	5	52	0	2	3	1	0	83.87%	82.54%	83.20%
DF	0	1	11	1	0	3	0	91.67%	68.75%	78.57%
MEL	1	3	0	82	10	35	0	72.57%	62.60%	67.21%
BKL	2	3	1	7	99	23	0	78.57%	73.33%	75.86%
NV	0	1	0	19	12	742	1	91.95%	95.74%	93.81%
VASC	0	0	0	0	0	1	17	**94.44%**	**94.44%**	**94.44%**
Class	**CF**									
	AKIEC	BCC	DF	MEL	BKL	NV	VASC	**Precision**	**Recall**	**F1-score**
AKIEC	33	1	0	0	3	2	0	**86.84%**	**80.49%**	**83.56%**
BCC	2	56	0	0	2	3	0	**94.92%**	**87.50%**	**91.08%**
DF	0	1	13	1	0	1	0	**92.86%**	**81.25%**	**86.67%**
MEL	0	0	0	88	7	35	1	**80.00%**	**67.18%**	**73.01%**
BKL	0	0	0	7	117	11	0	**93.60%**	**91.41%**	**92.49%**
NV	0	1	0	14	9	750	1	**93.51%**	**96.77%**	**95.11%**
VASC	0	1	0	0	0	0	17	89.47%	**94.44%**	91.89%

the specific context and requirements of the application. In scenarios where accurate diagnosis is critical, such as early detection of serious diseases, misdiagnosis can have severe consequences for patients. When the model is used for research purposes or in the development of treatment protocols, accuracy is generally the higher priority and considered the most important factor.

TABLE 6. Comparison of training time of each model and average prediction time for a dermoscopy image

Models	Training time	Prediction time for a dermoscopy image
ResNet-50	111m 44s	0.0111s
VGG-19	129m 40s	0.0139s
EfficientNet-V2	282m 19s	0.0111s
CF	152m 58s	0.0197s

5. **Conclusions.** This study explores the application of deep learning techniques to classify dermatoscopy images. We have successfully implemented and evaluated three individual models: ResNet-50, VGG-19, and EfficientNet-V2, most of which achieved positive results, however, EfficientNet-V2 achieved the best performance in classifying dermatoscopy images. To improve the classification performance and balance the performance between data classes, we improve by concatenating features from the three individual models into a new feature fusion model, which shows improved classification performance on all disease classes. This experiment's results demonstrate the potential application of the CF model in the field of medical imaging diagnosis, especially the diagnosis of skin diseases and skin cancer. With improvements in accuracy, the CF model opens up a new, promising direction for supporting effective and reliable medical diagnosis and decision-making.

Based on the findings from the research results, there are still challenges in classifying skin diseases from skin images. We have proposed improvement directions for future research:

- Continue to study more deeply the principles and mathematics of deep learning methods. While the current ensemble model has demonstrated promising results, combining additional models, including newer architectures like Vision Transformers (ViTs), could further enhance performance, particularly in complex cases or rare skin conditions.
- To address issues of class imbalance and enhance the model's ability to generalize, future work could explore more advanced data augmentation techniques. This could include transformations that simulate real-world variations in dermatological images, such as changes in lighting, angle, or noise.
- The model should be tested and validated on multiple and independent datasets. Search, collect more data sets, or create new data with high quality to conduct experiments and evaluate with different data sets to determine the applicability of the methods in each specific situation. In particular, it is necessary to add more data for labels (classes) with a low number of data units to ensure data balance.
- The CF model shows potential but rare classes such as MEL still present a challenge. Future improvements could involve techniques specifically designed to enhance performance on underrepresented classes, such as using techniques like class weighting, focal loss, or generative adversarial networks (GANs) for data synthesis.
- Study to improve the Ensemble model by improving the way to combine models with feature maps. Conduct experiments on different situations and contexts to clarify the strengths and limitations of the methodology.
- In medical applications, it is crucial for AI models to not only provide accurate predictions but also to offer explainable results. Future work could focus on enhancing the interpretability of the CF model's decisions, enabling dermatologists to understand why the model made a certain prediction and improving trust in its outputs.

By addressing these areas, future work can further push the boundaries of skin disease and cancer detection, creating even more accurate, efficient, and reliable AI-driven diagnostic tools.

Acknowledgment. This research is funded by University of Information Technology - Vietnam National University Ho Chi Minh City under grant number D1-2025-02.

REFERENCES

[1] World Health Organization. "Radiation: Ultraviolet (UV) radiation and skin cancer", *World Health Organization*, Retrieved October 16, 2017, website: https://www.who.int/news-room/questions-and-answers/item/radiation-ultraviolet-(uv)-radiation-and-skin-cancer

[2] Rebecca L. Siegel, Kimberly D. Miller, Ahmedin Jemal. "Cancer statistics, 2020", *CA: a cancer journal for clinicians*, 70(1), 2020, pp. 7-30, doi: 10.3322/caac.21590

[3] N. Mohanty, M. Pradhan, A. V. N. Reddy, S. Kumar, and A. Alkhayyat, "Integrated Design of Optimized Weighted Deep Feature Fusion Strategies for Skin Lesion Image Classification", *Cancers*, vol. 14, no. 22, 2022, doi: 10.3390/cancers14225716.

[4] X. Shen, L. Wei, and S. Tang, "Dermoscopic Image Classification Method Using an Ensemble of Fine-Tuned Convolutional Neural Networks", *Sensors*, vol. 22, no. 11, 2022, doi: 10.3390/s22114147

[5] H. Pehamberger, A. Steiner, and K. Wolff, "In vivo epiluminescence microscopy of pigmented skin lesions. I. Pattern analysis of pigmented skin lesions", *J. Am. Acad. Dermatol.*, vol. 17, no. 4, pp. 571–583, Oct. 1987, doi: 10.1016/S0190-9622(87)70239-4.

[6] Stolz. W, "ABCD rule of dermatoscopy: A new practical method for early recognition of malignant melanoma", *Eur J Dermatol*, vol. 4, pp. 521–527, 1994

[7] G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco, and M. Delfino, "Epiluminescence Microscopy for the Diagnosis of Doubtful Melanocytic Skin Lesions: Comparison of the ABCD Rule of Dermatoscopy and a New 7-Point Checklist Based on Pattern Analysis", *Arch. Dermatol.*, vol. 134, no. 12, pp. 1563–1570, Dec. 1998, doi: 10.1001/archderm.134.12.1563

[8] R. Garnavi, M. Aldeen, and J. Bailey, "Computer-Aided Diagnosis of Melanoma Using Border- and Wavelet-Based Texture Analysis", *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 6, pp. 1239–1252, Nov. 2012, doi: 10.1109/TITB.2012.2212282

[9] R. Sumithra, M. Suhil, and D. S. Guru, "Segmentation and Classification of Skin Lesions for Disease Diagnosis", *International Conference on Advanced Computing Technologies and Applications. ICACTA*, vol. 45, pp. 76–85, Jan. 2015, doi: 10.1016/j.procs.2015.03.090

[10] M. Q. Hatem, "Skin lesion classification system using a K-nearest neighbor algorithm", *Vis. Comput. Ind. Biomed. Art*, vol. 5, no. 1, p. 7, Mar. 2022, doi: 10.1186/s42492-022-00103-6

[11] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks", *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017, doi: 10.1038/nature21056

[12] J. Rathod, V. Waghmode, A. Sodha and P. Bhavathankar, "Diagnosis of skin diseases using Convolutional Neural Networks", *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, 2018, pp. 1048-1051, doi: 10.1109/ICECA.2018.8474593.

[13] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review", *Med. Image Anal.*, vol. 58, p. 101552, Dec. 2019, doi: 10.1016/j.media.2019.101552

[14] Gomathi, E. Jayasheela, M. Thamarai, M. Geetha, M.. (2023). "Skin cancer detection using dual optimization based deep learning network". *Biomedical Signal Processing and Control.* 84. 104968. 10.1016/j.bspc.2023.104968.

[15] K. Mridha, M. M. Uddin, J. Shin, S. Khadka and M. F. Mridha, "An Interpretable Skin Cancer Classification Using Optimized Convolutional Neural Network for a Smart Healthcare System", in *IEEE Access*, vol. 11, pp. 41003-41018, 2023, doi: 10.1109/ACCESS.2023.3269694.

[16] N. Gessert, M. Nielsen, M. Shaikh, R. Werner, A. Schlaefer, "Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data", *MethodsX*. 2020 Mar 19;7:100864, doi: 10.1016/j.mex.2020.100864. PMID: 32292713; PMCID: PMC7150512.

[17] M. A. Elashiri, A. Rajesh, S. N. Pandey, S. K. Shukla, S. Urooj, "Ensemble of weighted deep concatenated features for the skin disease classification model using modified long short term memory", *Biomedical Signal Processing and Control*, 76, 103729, 2022, doi: 10.1016/j.bspc.2022.103729

[18] Md. Minarul Islam Raju, Sumonto Sarker and Md. Mehedi Islam, "Chronic Kidney Disease Prediction using Ensemble Machine Learning", *Journal of Information Hiding and Multimedia Signal Processing*, Vol. 14, No. 1, pp. 1-9, March 2023.

[19] Ghada Abdallah Radwan, Mohamed Helmy Khafagy, Mostafa Thabet Mohamed Mabrouk, Marwa Hussien Mohamed, "Coronary Artery Disease Prediction by CombiningThree Classifiers", *Journal of Information Hiding and Multimedia Signal Processing*, Vol. 15, No. 4, pp. 221-235, September 2024.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition", *in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90

[21] Simonyan, K., and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", *3rd International Conference on Learning Representations (ICLR 2015)*, Computational and Biological Learning Society, 2015, pp. 1–14, doi: 10.48550/arXiv.1409.1556

[22] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks", *36th International Conference on Machine Learning, ICML 2019*, Long Beach, 9-15 June 2019, pp. 6105-6114, http://proceedings.mlr.press/v97/tan19a.html

[23] M. Tan and Q. Le, "EfficientNetV2: Smaller Models and Faster Training", *the 38th International Conference on Machine Learning*, PMLR, Jul. 2021, pp. 10096–10106, https://proceedings.mlr.press/v139/tan21a.html

[24] N. Codella, V. Rotemberg, P. Tschandl, M. E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti, H. Kittler, A Halpern, "Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC)", 2018; https://doi.org/10.48550/arXiv.1902.0336