IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 25 July 2024, accepted 20 September 2024, date of publication 26 September 2024, date of current version 10 October 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3468612

RESEARCH ARTICLE

EFAM-Net: A Multi-Class Skin Lesion Classification Model Utilizing Enhanced Feature Fusion and Attention Mechanisms

ZHANLIN JI[1,2,3], (Member, IEEE), XUAN WANG[1], CHUNLING LIU[4], ZHIWU WANG[4], NA YUAN[5], AND IVAN GANCHEV[3,6,7], (Senior Member, IEEE)

[1]College of Artificial Intelligence, North China University of Science and Technology, Tangshan 063210, China
[2]College of Mathematics and Computer Science, Zhejiang A&F University, Hangzhou 311300, China
[3]Telecommunications Research Centre (TRC), University of Limerick, Limerick, V94 T9PX Ireland
[4]Department of Pathology, Tangshan People's Hospital, Tangshan 063210, China
[5]Intelligence and Information Engineering College, Tangshan University, Tangshan 063000, China
[6]Department of Computer Systems, University of Plovdiv "Paisii Hilendarski", 4000 Plovdiv, Bulgaria
[7]Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, 1040 Sofia, Bulgaria

Corresponding authors: Zhiwu Wang (tcm2000@163.com), Na Yuan (yuanna@tsc.edu.cn), and Ivan Ganchev (ivan.ganchev@ul.ie)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFE0135700; in part by Bulgarian National Science Fund (BNSF) under Grant КП-06-ИП-КИТАЙ/1 (KP-06-IP-CHINA/1); and in part by the Telecommunications Research Centre (TRC), University of Limerick, Ireland.

ABSTRACT Skin cancer caused by common malignant tumors is a major threat to the health of patients. Automated classification of skin lesions using computer algorithms is crucial for enhancing diagnostic efficiency and reducing mortality rates associated with skin cancer. Enhancing the capabilities of image classification models for skin lesions is essential to assist in accurately classifying skin diseases of patients. Aiming at this goal, a novel EFAM-Net model is proposed in this paper for the skin lesion classification task. Firstly, a newly designed Attention Residual Learning ConvNeXt (ARLC) block is used to extract low-level features such as colors and textures in images. Then, the deep-layer blocks of the network are replaced with a newly designed Parallel ConvNeXt (PCNXt) block, allowing to capture richer and more complex features. Additionally, another newly designed Multi-scale Efficient Attention Feature Fusion (MEAFF) block enhances feature extraction at various scales, allowing the model to effectively capture more comprehensive features in specific layers, fuse feature maps of different scales and enhance feature reuse at the end. EFAM-Net is experimentally evaluated on the ISIC 2019 and HAM10000 public datasets, as well as on a private dataset. The obtained results show that EFAM-Net achieves top classification performance among all compared models, by achieving overall accuracy of 92.30%, 93.95%, and 94.31% on the ISIC 2019, HAM10000, and private dataset, respectively.

INDEX TERMS Skin cancer, lesion classification, ConvNeXt, EFAM-Net, ISIC 2019, HAM10000.

I. INTRODUCTION

Skin cancer is a common cancer that affects a substantial number of people worldwide. In the United States alone, skin cancer affects 3 million people annually, [1]. Skin cancer is a dangerous form of cancer that can be classified into three primary classes: melanoma, basal cell carcinoma, and squamous cell carcinoma in situ, [2]. Among them, melanoma is particularly deadly, arising from the abnormal proliferation and mutation of melanocytes. Prompt treatment of melanoma can significantly impact patient survival rates, despite the high mortality rate associated with advanced melanoma. Therefore, early detection and diagnosis are crucial for improving patient survival rates.

In the past, the diagnosis of skin cancer was mainly performed through non-auxiliary methods such as hospital clinical examinations, due to the subjectivity of visual analysis even with long time spent for human training, [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Rajeswari Sundararajan.

Therefore, the accuracy of these diagnostic methods is limited and there is uncertainty, which brings great challenges to the detection of skin diseases as early as possible. In addition, clinical examinations require medical professionals to spend lots of time on checking the patients' conditions, which can delay treatment and cause skin conditions to worsen, leading to the consumption of more medical resources.

Imaging technology has advanced rapidly, and more and more medical professionals are using dermoscopy to diagnose skin cancer in patients, [4]. Dermoscopy is a non-reflective, non-invasive imaging method that magnifies images of skin lesions up to ten times, [5].

In addition, the continuous development of computer technology and improvement in hardware performance enabled artificial intelligence (AI)-based techniques to make noteworthy progress in the field of image processing. Among these, convolutional neural networks (CNNs) play a key role in helping doctors to automatically and efficiently extract key features of skin lesions, through a large number of trainings on skin lesion image datasets, for early detection and diagnosis of skin cancer.

Although CNNs have achieved good results in assisting doctors in examining skin diseases, these networks still face the following challenges: (1) in dermoscopic images, the differences in features between normal skin and lesion areas are minor and, in addition, artifacts could be present, which makes problematic the features extraction; (2) some classes of skin lesions have very similar color, texture, and lesion appearance to each other; and (3) some other classes of skin lesions have high intra-class variability. All these make automatic classification difficult to perform. This paper proposes a new network model for the classification of multi-class skin lesions with the following main contributions:

1) A newly designed Attention Residual Learning ConvNeXt (ARLC) block is proposed for use in the shallow network to improve the attention learning mechanism of the original ConvNeXt block, allowing the proposed model to capture low-level image features such as colors and textures.
2) To enhance feature propagation and reuse, a newly designed Parallel ConvNeXt (PCNXt) block is proposed to obtain richer and more complex image features by extracting global information, which further enhances the classification abilities of the proposed EFAM-Net model.
3) To fuse features at various levels effectively, a newly designed Multi-scale Efficient Attention Feature Fusion (MEAFF) block is proposed, which uses shallow features to complement and enrich deep features, thus effectively solving problems associated with small differences between features.

The rest of the paper is structured as follows. Section II presents related works on skin lesion classification. Section III describes the proposed EFAM-Net model in detail. Section IV introduces the experimental setting parameters, the conducted ablation study and performance comparison experiments, and analyzes the obtained results. Finally, Section V concludes the paper and identifies future research directions.

II. RELATED WORK

A. CLASSIFICATION OF SKIN LESIONS BASED ON DEEP LEARNING

Deep learning is a computer-aided technique with minimal requirements for manual feature engineering. CNNs are one of the most commonly used deep learning models with the ability to extract information and learn. Through training, CNNs can accurately classify skin lesions. For instance, Gilani et al. [6] proposed a novel model to identify skin malignancies using deep spiking neural networks with higher efficiency compared to traditional deep neural networks. Specifically, these authors proposed a spiking model, based on VGG-13 [7], which uses impulse neurons instead of traditional neurons. The model achieved 89.57% accuracy and 90.07% F1 score on the ISIC 2019 dataset. Villa-Pulgarin et al. [8] conducted four experiments on skin lesion classification using three commonly used deep learning models, namely DenseNet-201 [9], Inception-ResNet-V2 [10], and Inception-V3 [11]. Their experiments demonstrated the effectiveness of deep learning and transfer learning in skin lesion recognition. Skin lesion classification methods based on deep learning have made considerable progress, exhibiting high classification performance. Zeng et al. [12] proposed an enhanced multi-source knowledge fusion distillation framework, which improves knowledge transfer in a dual-stage progressive distillation approach to maximize mutual information between teacher and student representations. The deep learning model using this distillation method achieved excellent performance on multiple skin lesion datasets and demonstrated remarkable effectiveness with a smaller computational burden. However, these methods suffer from poor interpretability, making it challenging to explain the basis on which a model classifies images.

B. APPLICATION OF FEATURE FUSION IN SKIN LESION CLASSIFICATION

Rich features can provide more comprehensive information for a model, and many researchers have found that fusing different feature information can indeed improve the model classification performance. For instance, Maurya et al. [13] proposed a transfer learning-based feature fusion approach for benign and malignant melanoma identification. They used different deep CNNs to extract features and then applied an XGBoost-based feature selection method to eliminate useless features and perform feature fusion. The selected features are used to train and test the XGBoost classifier. Experimental results show that their proposed feature fusion and selection strategy achieves higher accuracy compared to others. Mahbod et al. [14] proposed a fusion method, based on multiple CNNs, whereby EfficientNet B0 [15],

EfficientNet B1, and SeResNeXt-50 [16] were used as backbone models to extract features from cropped images at six different scales, which are then fused for image classification. Maqsood and Damasevicius [17] proposed a framework for multi-class skin cancer segmentation and classification, using modified Xception [18], ResNet-50 [19], ResNet-101 [19], and VGG16 [7] for feature extraction. The Convolutional Sparse Image Decomposition (CSID) algorithm was used to extract selected feature vectors, which are then combined into a matrix for classification purposes. Finally, a multi-class support vector machine (SVM) classifier was employed for classification. Previous research has often focused on extracting features for subsequent fusion by employing multiple CNNs. However, features extracted from multiple CNNs of similar depth tend to be close and similar to each other, which results in an increased number of model parameters and reduced efficiency.

C. ATTENTION-MECHANISM-BASED SKIN LESION CLASSIFICATION

Appropriate application of attention mechanisms can enhance the model's focus on lesion regions, thereby increasing the efficiency of feature extraction. For example, Lan et al. [20] proposed an improved capsule network with an incorporated Convolutional Block Attention Module (CBAM) [21] to address the issue of spatial information loss. The improved model performs excellently for early detection of skin cancer. Karthik et al. [22] replaced the original Squeeze-and-Excitation block in the EfficientNetV2 [23] with the Efficient Channel Attention (ECA) block [24], achieving 84.70% accuracy on a dataset containing images of four skin lesion classes – acne, actinic keratosis, melanoma, and psoriasis. Liu et al. [25] proposed a feature fusion CNN strategy, based on a multi-branch attention, which combines a linear bottleneck inverted residual block with a local cross-channel attention to construct an inverted residual block with channel attention. This method allows to extract more comprehensive features, leading to an improvement in classification accuracy. Nevertheless, the introduction of attention mechanisms necessitates learning a substantial number of additional parameters, which may not only incur additional computational costs but can also potentially lead to model overfitting, especially when dealing with small training datasets.

III. PROPOSED MODEL: EFAM-Net

A. OVERALL STRUCTURE

The proposed EFAM-Net model is based on the ConvNeXt network [26] and consists of a feature extraction layer, a feature fusion layer, and a linear layer, as depicted in Figure 1. There are four stages in the feature extraction layer, whereby stage 2 includes nine blocks and the other three stages include three blocks. The ConvNeXt block in stage 0 is replaced by a newly designed ARLC block to help the shallow network improve its discrimination ability. The ConvNeXt block in stage 3 is replaced by a newly designed PCNXt block, which improves the model's learning ability for deep semantic information so that the gradient can better propagate to the shallower network. A newly developed feature fusion strategy is introduced into the model's feature fusion layer, which fuses multi-scale features, enhances the dependency between local and global features, and enriches feature information to further improve the classification performance of the model. Finally, the feature information is normalized and predicted through the linear layer.

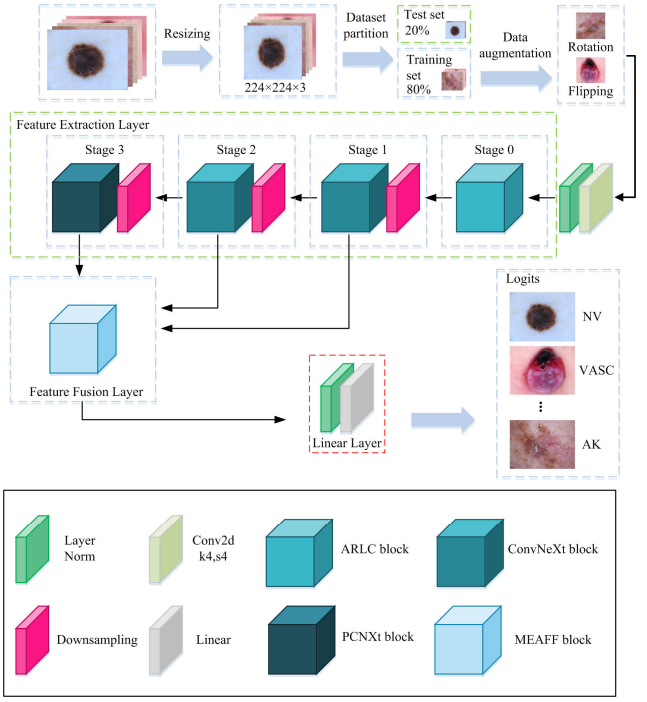


FIGURE 1. The proposed EFAM-Net model.

To improve the robustness and generalization ability of the proposed EFAM-Net model, dataset preprocessing and data augmentation operations [30] were performed without changing the number of images, as follows:

1) The original images were resized to 224 × 224 pixels to ensure the consistency of the input size and assist the model in concentrating on critical regions.
2) On the training set, data augmentation operations were performed, such as random rotation, random horizontal (left/right) flipping, and histogram equalization, as to increase the diversity of the image information and alleviate overfitting during the model training process.

B. ARLC BLOCK

Due to the complex features of skin lesion images, such as texture and color, it is difficult for the original ConvNeXt block to extract rich contextual information. Inspired by Zhang et al. [27], a newly designed block is presented here, called ARLC (Figure 2). The ARLC block generates attention

weights internally and implements residual attention learning to improve the representation ability of the model. It effectively mitigates issues such as model overfitting and gradient explosion that may arise from directly incorporating the attention module into the block.

In the design of the ARLC block, we combined a depthwise convolution, a layer normalization, and a 1×1 convolution to form a depthwise separable convolution, which can extract more local features and channel information with less computational cost. Based on this, we implemented an attention mechanism using the residual network approach.

Initially, the ARLC takes the input feature map $X_i \in \mathbb{R}^{H \times W \times C}$ in the format of a particular height (H) $\times$ width (W) $\times$ number of channels (C). The input feature map is fed into the three branches of ARLC, whereby the first branch is used for identity mapping, and the other two branches start with depthwise convolution. The general operation of these three branches is performed according to the following formula:

$$Branch_l^i = \begin{cases} X_i; & l=1 \\ Conv_{1\times1}(GELU\ (Conv_{1\times1}\ (LN(DWSC\ (X_i))))); & l=2 \\ X_i \otimes \sigma(LN(Branch_2^i)) \otimes \alpha; & l=3 \end{cases} \tag{1}$$

where i denotes the index of the ARLC block, $Branch_l^i$ denotes the output of the l-th branch, $Conv_{1\times1}$ denotes an 1×1 convolution, $DWSC$ denotes a depthwise convolution, $\otimes$ denotes an element-wise multiplication, $GELU$ denotes the Gaussian Error Linear Units (GELU) activation, α is a learnable weighting factor, LN denotes a layer normalization, and σ denotes the SoftMax function.

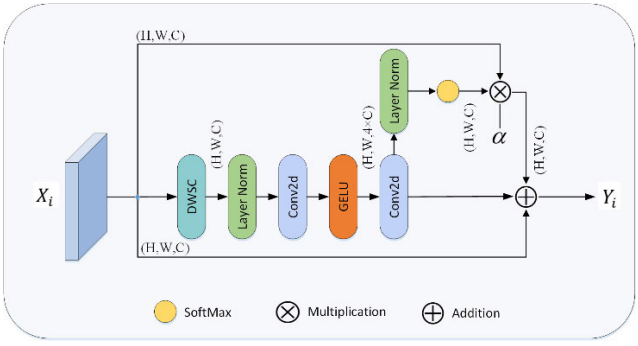


FIGURE 2. The newly designed ARLC block, utilized by the proposed EFAM-Net model.

In the third branch, first a layer normalization is applied to $Branch_2^i$ in the feature channel dimension, and then the SoftMax function is used to weight the features, so that the model can pay more attention to the important regions in each channel, thereby generating an attention mask. Similarly to the gate usage in highway networks [28], the attention mask is used as a control gate of X_i, and the attention feature map is obtained by an element-wise multiplication. This attention mechanism enhances the ability of the model to focus on important semantic regions without the need for additional parameters. The learnable weighting factor α is utilized to adaptively adjust the contribution of the attention feature map in order to avoid introducing the noise of the early feature map into the deep features. The $Branch_3^i$ maintains the resistance to noise and enhances the locating information.

The output feature Y_i is obtained by ARLC as follows:

$$Y_i = Branch_1^i \oplus Branch_2^i \oplus Branch_3^i \tag{2}$$

where $\oplus$ denotes an element-wise addition.

C. PCNXt BLOCK

The deepness of a network is crucial in enabling the extraction of complex features as it allows to capture advanced and abstract information of images. To improve the model's learning ability without increasing the number of parameters, we designed the PCNXt block, shown in Figure 3. The PCNXt block introduces a parallel branch with a pooling layer and layer normalization. The pooling layer can extract global information of features with higher semantic and generalization ability. In addition, it helps extract background information to identify the type and location of skin lesions. The layer normalization can standardize the feature map and enhance features while reducing the number of parameters and computational complexity of the model.

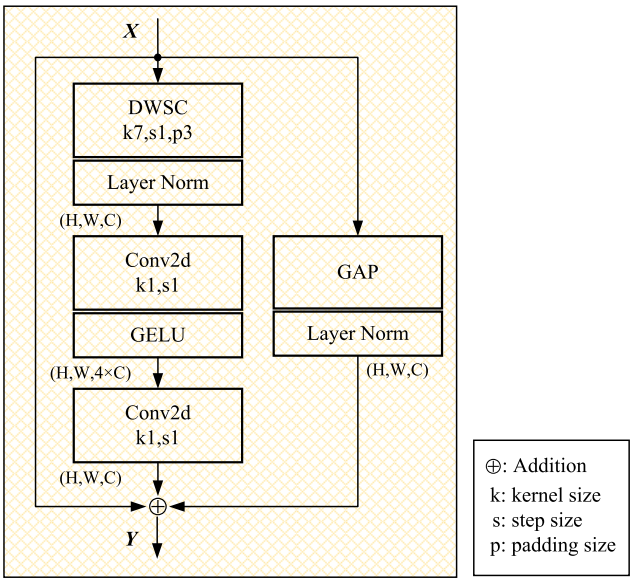


FIGURE 3. The newly designed PCNXt block, utilized by the proposed EFAM-Net model.

The output Y of the PCNXt block is obtained by an element-wise addition of the identity map, residual feature map, and global feature map, as follows:

$$Y = X \oplus Conv_{1\times1}\left(GELU(Conv_{1\times1}\ ((LN\ (DWSC\ (X)))))\right) \oplus LN(GAP(X)) \tag{3}$$

where X denotes the input feature map, $DWSC$ denotes a depthwise convolution, and GAP denotes a global average pooling.

The newly designed PCNXt block replaces the ConvNeXt block only in stage 3 of the feature extraction layer of the proposed model. If the blocks in other stages were replaced by PCNXt blocks, it would only have increased the model complexity with a little improvement to its classification performance.

D. MEAFF BLOCK

We designed a new feature fusion block, named MEAFF, to fuse the feature maps between different layers in order to refine and enrich the feature information. The features extracted by the shallow network contain more positional information, which is advantageous for the model to acquire information such as the location and texture of skin lesions. The features extracted by the deep network contain more crucial semantic information for identifying different skin lesions. Before fusing multi-level features, we use the Efficient Channel Attention (ECA) block to generate corresponding weight values based on the importance of each channel in the feature map, enabling the model to focus more on important feature channels and enhancing its representative capacity for fused information. The detailed design of the MEAFF block structure is shown in Figure 4.

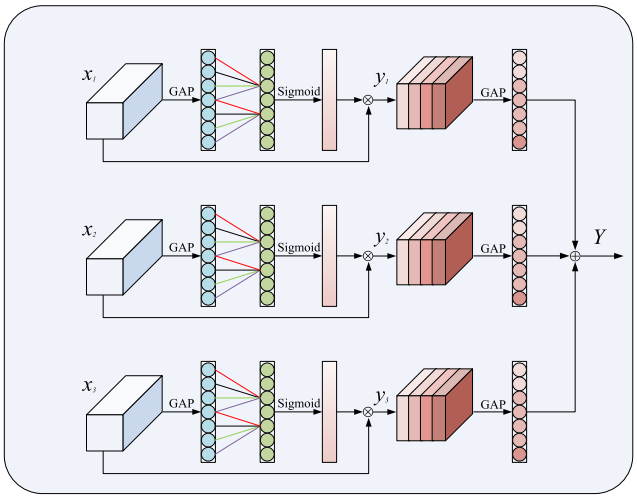


FIGURE 4. The newly designed MEAFF block, utilized by the proposed EFAM-Net model.

The input to the MEAFF block comes from feature maps of Stages 1−3 of the feature extraction layer. Initially, channel-level weights are assigned to each of the three feature maps layer-by-layer to obtain sufficiently effective feature information. This addresses the issue of ambiguity or loss of lesion location and features. Then, the following operation is performed:

$$y_i = \sigma(Conv_{k\times k}(GAP(x_i))) \otimes x_i \tag{4}$$

where $x_i \in \mathbb{R}^{H\times W\times C}$ $(i = \{1, 2, 3\})$ represents the feature map of the i-th stage of the feature extraction layer and $Conv_{k\times k}$ denotes a cross-channel one-dimensional convolution with a kernel size k, defined as:

$$k = \left|\frac{log_2 C}{\gamma} + \frac{b}{\gamma}\right|_{odd} \tag{5}$$

where $|x|_{odd}$ denotes the nearest odd number greater than or equal to x, γ and b are hyperparameters, and C denotes the number of channels of the input feature map.

There are three different down-sampling operations in the feature extraction layer of EFAM-Net, so the three feature maps of the corresponding stages have different channel sizes. Therefore, a global average pooling (GAP) is used to compress the three feature maps to achieve channel size unity. Finally, the three feature maps are merged through an element-wise addition to generate a feature map, containing strong semantic information while preserving the details of the original feature maps, as follows:

$$Y = GAP(y_1) \oplus GAP(y_2) \oplus GAP(y_3) \tag{6}$$

The MEAFF block improves lesion localization and overcomes difficulties in recognizing similar textures by following a multi-scale attention feature fusion strategy.

E. PREDICTION AND OPTIMIZATION

After obtaining logits from the model's linear layer, the SoftMax activation function normalizes them. The SoftMax activation function converts unnormalized values into a probability distribution, ensuring that each class is associated with a corresponding probability value and the sum of probabilities of all classes equals 1. This step achieves the transformation of the model's output into a probabilistic interpretation, facilitating more precise decisions among various classes by the model. The SoftMax activation function is defined as follows:

$$\sigma\left(x_j\right) = \frac{e^{x_j}}{\sum_{i=1}^{n} e^{x_i}} \tag{7}$$

where $X = (x_1, \ldots, x_n) \in \mathbb{R}^n$ denotes logits.

In addition, the cross-entropy loss function, commonly used in classification tasks, is utilized to measure the discrepancy between the model's predicted results and the true labels. Once the class probability distribution is obtained with the SoftMax activation function, it is used to calculate the distance from the true probability distribution using cross entropy. This distance is an important basis for calculating the gradient. Finally, the model's parameters are updated through a gradient back propagation to continuously optimize the model.

The multi-class cross-entropy loss function value is calculated as follows:

$$L = -\sum_{i=1}^{n} t_i log(p_i) \tag{8}$$

where t_i denotes the true label for the i-th class and p_i denotes the model's predicted probability for the i-th class.

IV. EXPERIMENTS AND RESULTS

A. DATASETS

Experiments were conducted on the ISIC 2019 [29], [30] and HAM10000 [31] public datasets, which are freely available from the International Skin Imaging Collaboration Challenge, and a private dataset. The ISIC 2019 dataset includes 25,331 labeled images, classified into eight classes: actinic keratosis (AK), basal cell carcinoma (BCC), benign keratosis-like lesions (BKL), dermatofibroma (DF), melanoma (MEL), melanocytic nevus (NV), squamous cell carcinoma (SCC), and vascular skin lesions (VASC). The HAM10000 dataset includes 10,015 images of skin lesions, classified into seven classes: actinic keratosis and intraepithelial carcinoma (AKIEC), BCC, BKL, DF, MEL, NV, and VASC. The private dataset was provided by Peking Union Medical College Hospital. It includes six classes of skin lesions, totaling 2,900 images, as follows: 1615 images of acne (AC), 125 images of discoid lupus erythematosus (DLE), 400 images of melasma (ME), 280 images of Ota nevus (ON), 310 images of rosacea (ROS), and 170 images of seborrheic dermatitis (SD) skin lesions. In addition, to ensure the authenticity of the experimental results and reduce the interference caused by errors in the experimental data, all images in this dataset were strictly reviewed by a number of experienced dermatologists. As with the public datasets, three types of data augmentation operations (i.e., random rotation, random horizontal (left/right) flipping, and histogram equalization) were performed on the training set before training the model. Sample images of various skin lesions, contained in this dataset, are shown in Figure 5.

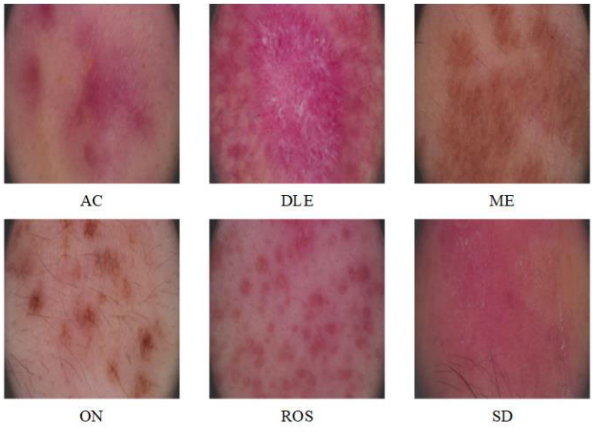


FIGURE 5. Sample images of the private dataset.

Among the three datasets, ISIC 2019 is the largest one, containing more classes than the other two datasets, but suffering from class imbalance issues. HAM10000 exhibits significant intra-class sample differences and minor inter-class sample differences. In contrast, the private dataset is the smallest in size, posing challenges for the model training.

TABLE 1. Splitting the datasets into training and test sets.

Datasets	Classes	Total number of images	Number of images in training set	Number of images in test set
ISIC 2019	AK	867	694	173
	BCC	3323	2659	664
	BKL	2624	2100	524
	DF	239	192	47
	MEL	4522	3618	904
	NV	12875	10300	2575
	SCC	628	503	125
	VASC	253	203	50
HAM10000	AKIEC	327	262	65
	BCC	514	412	102
	BKL	1099	880	219
	DF	115	92	23
	MEL	1113	891	222
	NV	6705	5364	1341
	VASC	142	114	28
Private dataset	AC	1615	1292	323
	DLE	125	100	25
	ME	400	320	80
	ON	280	224	56
	ROS	310	248	62
	SD	170	136	34

To ensure fairness and rigor of experiments, we randomly divided each dataset into a training set and a test set using a ratio of 8:2. Table 1 shows the number of images in every class of each dataset, before and after splitting it.

B. EXPERIMENTAL SETUP

We employed the PyTorch deep learning framework to construct our network model. The programming language utilized was Python 3.8, operating on a Linux system. The hardware setup utilized an NVIDIA GeForce RTX 3090 GPU.

During the model training, we meticulously controlled each hyperparameter, ensuring consistent hyperparameter initialization across all models that participated in the experiments. The optimization was performed using the Adaptive Moment Estimation with Weight Decay (AdamW) optimizer [32]. The specific hyperparameter values used are shown in Table 2. Additionally, all models were pre-trained on ImageNet [33] and initialized with the pre-trained parameters.

C. EVALUATION METRICS

To evaluate the classification performance of the compared models, we used accuracy, precision, recall, specificity,

TABLE 2. Hyperparameter values.

Hyperparameters	Settings
Initial learning rate	5e-4
Minimum learning rate	5e-10
Batch size	64
Epochs	40
Loss	Cross-entropy loss function
Optimizer	AdamW
Learning rate optimization strategy	Cosine annealing strategy

and F1 score, calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

$$Specificity = \frac{TN}{TN + FP} \tag{12}$$

$$F1_{score} = \frac{2 \times precision \times recall}{precision + recall} \tag{13}$$

where TP denotes the true positive count, TN denotes the true negative count, FP denotes the false positive count, and FN denotes the false negative count.

In addition, we used also the macro-averaged accuracy, treating each class equally by assigning the same weight, which is calculated by taking the arithmetic mean of accuracies across all classes, as follows:

$$Macro\text{-}averaged\ accuracy = \frac{1}{n}\sum_{i=1}^{n} Accuracy_i \tag{14}$$

where n denotes the number of classes, and $Accuracy_i$ denotes the accuracy for class i.

D. ABLATION STUDY EXPERIMENTS

To evaluate the effectiveness of each newly designed block type on the performance of the proposed EFAM-Net model, we conducted ablation study experiments on the ISIC 2019 dataset, using ConvNeXt as a baseline. Table 3 provides the details of various component combinations resulting in different model versions (the best results are shown in **bold**).

In the first set of experiments, the effects of the newly designed blocks on the proposed EFAM-Net model were individually evaluated. These experiments demonstrated that all three blocks contributed positively to improving the classification performance of the developed model, with the highest contribution given by the MEAFF block. In the second set of experiments, the newly designed blocks were respectively added in pairs to the baseline. Among these, the combination of the PCNXt block and MEAFF block provided the best effect according to three out of five metrics, followed closely by the 'ARLC + MEAFF' combination, which indicates that the collocation of attention mechanisms can capture the correlation between features more efficiently. In the final experiment, the three newly designed block types were added together to the baseline to form the proposed EFAM-Net model, which resulted in improvements in all metrics, except for average recall. This indicates that integrating all these block types helps the model learn richer feature information from skin lesion images.

TABLE 3. Ablation study results, obtained on the ISIC 2019 dataset.

Models	Evaluation metrics (%)				
	Overall accuracy	Average recall	Average precision	Average specificity	Average F1 score
ConvNeXt (*baseline*)	90.91	88.69	84.92	98.32	86.66
Baseline + ARLC	91.11	89.37	86.28	98.33	87.63
Baseline + PCNXt	91.18	88.78	86.18	98.33	87.40
Baseline + MEAFF	91.88	**89.80**	87.11	98.46	88.35
Baseline + ARLC + PCNXt	91.31	89.07	86.14	98.36	87.48
Baseline + ARLC + MEAFF	92.20	88.70	86.60	98.55	87.64
Baseline + PCNXt + MEAFF	91.96	89.35	87.43	98.50	88.32
Baseline + ARLC + PCNXt + MEAFF (i.e., **EFAM-Net *proposed***)	**92.30**	88.86	**88.27**	**98.59**	**88.52**

E. CLASSIFICATION PERFORMANCE COMPARISON

To compare the classification performance of the proposed EFAM-Net model to other existing models, experiments were conducted on the three datasets. Figures 6 and 7 show the classification performance of the proposed model on the ISIC2019 and HAM10000 public datasets, respectively, in the form of confusion matrices [34]. There, the accuracy of predicting each skin lesion class is represented by the diagonal values, while the color intensity of the squares indicates the level of accuracy in predicting each class. As can be seen in Figure 6, the proposed EFAM-Net model performs particularly well in the classification of classes BCC, NV, and VASC, which contain considerable number of images. Moreover, the accuracy of the proposed EFAM-Net model in all classes exceeds 80%, which is a demonstration of a particularly good overall classification performance. Similar conclusion can be drawn from Figure 7.

Table 4 shows classification performance comparison of EFAM-Net with baseline models. As can be seen, EFAM-Net outperforms all of them according to all evaluation metrics

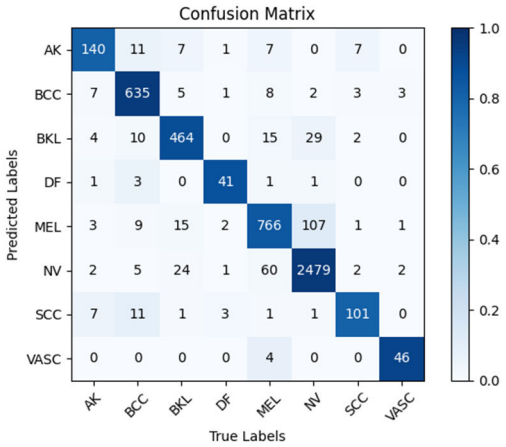


FIGURE 6. The confusion matrix of the proposed EFAM-Net model on the ISIC2019 dataset.

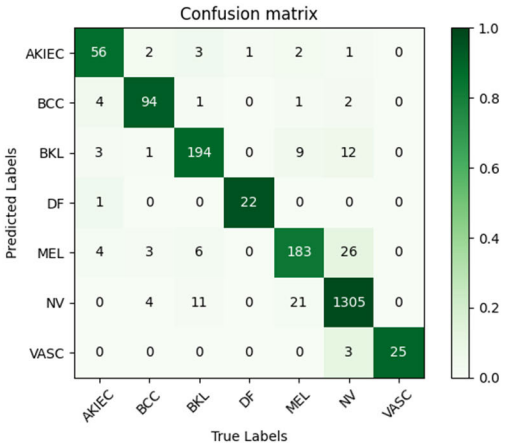


FIGURE 7. The confusion matrix of the proposed model on the HAM10000 dataset.

across all three datasets (except for average precision on HAM10000 where it gives way to first place to ResNet-101), indicating that the proposed model can classify skin lesions more correctly, which may have profound implications for practical applications.

Figure 8 shows the accuracy and loss curves of the proposed EFAM-Net model, in comparison with the baseline (ConvNeXt), on the ISIC 2019 dataset. The ConvNeXt performance reaches its maximum around the 25th training epoch, with signs of overfitting, while the curves of the EFAM-Net model gradually level off after the 35th epoch, reaching maximum performance. Although the EFAM-Net model utilizes a newly designed ARLC block with its own production of attention weights, and a newly designed MEAFF block with a lightweight attention mechanism, this has an insignificant impact on the model complexity and does not introduce a risk of overfitting. Besides, the number of parameters of the EFAM-Net model (28.8M) is only about 3.6% more than that of the baseline model (27.8M), which also has a minor impact on the model complexity.

To evaluate the multi-class skin lesion classification abilities of the proposed model more comprehensively, we compared its accuracy achieved in each skin lesion class to that of the baseline models on the two public datasets. As can be seen in Table 5, on the ISIC 2019 dataset, EFAM-Net

TABLE 4. Classification performance comparison of the proposed EFAM-Net model with baseline models.

Datasets	Models	Evaluation metrics (%)				
		Overall accuracy	Average recall	Average precision	Average specificity	Average F1 score
ISIC 2019	ResNet-101	87.48	82.80	79.55	97.74	81.03
	DenseNet-201	89.13	86.29	83.06	98.00	84.57
	EfficientNet B0	89.83	86.14	83.54	98.11	84.75
	ConvNeXt	90.91	88.69	84.92	98.32	86.66
	EFAM-Net (*proposed*)	**92.30**	**88.86**	**88.27**	**98.59**	**88.52**
HAM10000	ResNet-101	89.35	84.96	**94.99**	97.37	84.95
	DenseNet-201	90.80	86.80	86.03	97.46	86.01
	EfficientNet B0	91.60	87.28	84.71	97.62	85.91
	ConvNeXt	92.35	87.64	88.45	97.95	87.93
	EFAM-Net (*proposed*)	**93.95**	**91.44**	90.23	**98.44**	**90.78**
Private dataset	ResNet-101	91.72	88.28	82.63	97.84	84.82
	DenseNet-201	92.76	89.01	86.01	98.14	87.31
	EfficientNet B0	90.86	87.89	83.52	97.71	85.15
	ConvNeXt	92.93	90.19	87.11	98.19	88.51
	EFAM-Net (*proposed*)	**94.31**	**92.61**	**88.74**	**98.49**	**90.25**

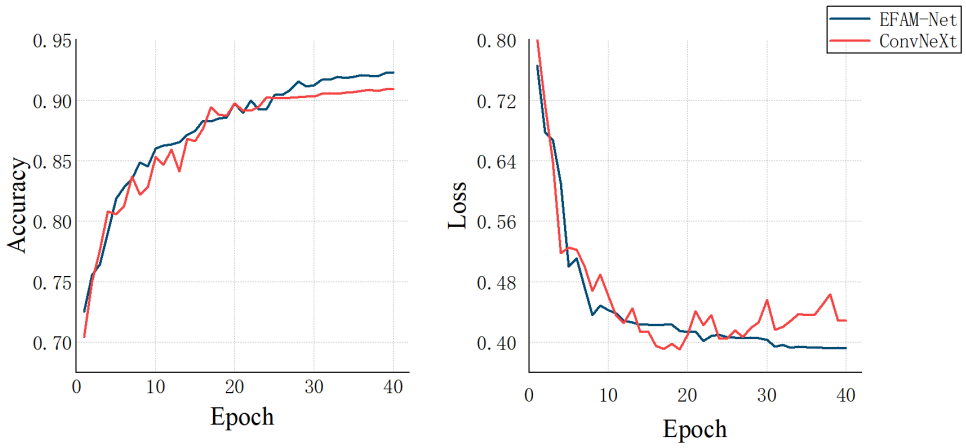


FIGURE 8. The accuracy and loss curves of the proposed EFAM-Net model vs. ConvNeXt on the ISIC 2019 dataset.

TABLE 5. Multi-class classification performance comparison of the proposed EFAM-Net model with baseline models w.r.t. accuracy (%), achieved on the ISIC 2019 dataset.

Models	Class accuracy								Macro-averaged accuracy
	AK	BCC	BKL	DF	MEL	NV	SCC	VASC	
ResNet-101	98.08	97.31	95.75	99.55	92.75	92.89	98.91	99.66	96.86
DenseNet-201	98.36	97.55	96.46	99.74	93.97	93.54	98.83	**99.80**	97.28
EfficientNet B0	98.40	97.95	96.62	99.74	94.51	93.99	98.78	99.78	97.47
ConvNeXt	98.79	98.08	96.88	**99.76**	94.98	94.57	98.95	**99.80**	97.73
EFAM-Net (*proposed*)	**98.87**	**98.46**	**97.79**	99.72	**95.38**	**95.34**	**99.23**	**99.80**	**98.07**

TABLE 6. Multi-class classification performance comparison of the proposed EFAM-Net model with baseline models w.r.t. accuracy (%), achieved on the HAM10000 dataset.

Models	Class accuracy							Macro-averaged accuracy
	AKIEC	BCC	BKL	DF	MEL	NV	VASC	
ResNet-101	98.70	98.75	95.60	99.70	93.80	92.35	99.80	96.96
DenseNet-201	98.75	98.60	96.10	99.75	94.80	93.75	99.85	97.37
EfficientNet B0	98.65	98.85	96.40	99.70	95.45	94.55	99.80	97.63
ConvNeXt	98.60	98.90	96.90	99.75	95.65	95.00	**99.90**	97.81
EFAM-Net (*proposed*)	**98.95**	**99.10**	**97.70**	**99.90**	**96.40**	**96.00**	99.85	**98.27**

achieved the highest classification accuracy in seven skin lesion classes, except for BCC. On the HAM10000 dataset, EFAM-Net achieved the highest accuracy in six skin lesion classes, except for VASC, as shown in Table 6. Additionally, EFAM-Net outperforms all baseline models w.r.t. the macro-averaged accuracy on both datasets.

Next, we compared the classification performance of the proposed EFAM-Net model with that of state-of-the-art models as reported in the literature for the two public datasets. The obtained results are shown in Tables 7 and 8 ("-" denotes missing data in the source). As can be seen in Table 7, on the ISIC 2019 dataset, EFAM-Net surpasses all other models

TABLE 7. Classification performance and computational complexity comparison of the proposed EFAM-Net model with state-of-the-art models on the ISIC 2019 dataset.

Models	Evaluation metrics (%)					Parameter count (M)
	Overall accuracy	Average recall	Average precision	Average specificity	Average F1 score	
Xception [36]	88.00	88.00	88.00	-	88.00	-
MCELM [37]	89.00	88.25	88.25	-	-	-
Cubic SVM [38]	91.71	84.82	**92.04**	96.40	86.82	-
SSD-KD [39]	86.60	-	-	-	-	66.7
CLCM-NET [40]	91.73	85.30	84.80	-	85.05	-
Hybrid Model [41]	91.93	-	85.58	98.29	-	-
Swin Transformer–based model [42]	89.36	88.22	85.13	-	86.65	-
EFAM-Net (*proposed*)	**92.30**	**88.86**	88.27	**98.59**	**88.52**	**28.8**

TABLE 8. Classification performance and computational complexity comparison of the proposed EFAM-Net model with state-of-the-art models on the HAM10000 dataset.

Models	Evaluation metrics (%)					Parameter count (M)
	Overall accuracy	Average recall	Average precision	Average specificity	Average F1 score	
RegNetY [43]	91.00	-	-	-	88.10	-
EW-FCM [44]	86.33	-	86.33	97.72	-	**1.8**
Fused CNNs [45]	90.85	83.75	83.81	-	83.45	-
G-DMN [46]	87.07	80.13	75.28	96.19	77.26	5.3
EL-DLOA [47]	88.80	83.70	89.70	-	86.20	-
Enhanced ShuffleNet [48]	89.31	89.54	87.78	-	88.65	-
FFN [49]	92.00	75.00	-	-	76.00	-
EFAM-Net (*proposed*)	**93.95**	**91.44**	**90.23**	**98.44**	**90.78**	28.8

according to four (out of five) evaluation metrics. The only exception is the average precision, where EFAM-Net takes second place after the winner (Cubic SVM). More specifically for all other four metrics, the second-best performer was left behind by 0.37 percentage points (w.r.t. Hybrid Model) according to overall accuracy, 0.61 percentage points (w.r.t. MCELM) based on average recall, 0.30 percentage points (w.r.t. Hybrid Model) according to average specificity, and 0.52 percentage points (w.r.t. Xception) based on average F1 score. On the HAM10000 dataset (c.f., Table 8), EFAM-Net is the absolute winner, according to all evaluation metrics. These results demonstrate the effectiveness of EFAM-Net in performing skin lesion classification tasks.

F. VISUALIZATION OF CLASSIFICATION RESULTS

Finally, we applied the Gradient Weighted Class Activation Mapping (Grad-CAM) [42] Explainable Artificial Intelligence (XAI) technique for visual analysis of the classification results of the proposed EFAM-Net model compared to the baseline (ConvNeXt). We randomly selected a sample from each of the eight different classes of lesion images, contained in the ISIC 2019 dataset, and compared the classification results of the baseline and the proposed EFAM-Net model, as depicted in Figure 9.

This visualization demonstrates the model's attention distribution across various regions in the images, with warmer colors indicating higher levels of attention. Compared with ConvNeXt, the proposed EFAM-Net model demonstrates higher attentional capacity to focus on the lesion portion, rather than on the background tissue or other noisy information. More specifically, in the images of AK (1st column) and VASC (8th column), ConvNeXt erroneously focuses on noise information. In the images of MEL (5th column) and NV (6th column), ConvNeXt incorrectly attends to normal skin areas. In contrast, EFAM-Net demonstrates more accurate attention distribution, significantly concentrating on the regions of skin lesions. The proposed model's ability to focus on semantically meaningful aspects of lesions is primarily attributed to the newly designed blocks. These blocks judiciously apply feature fusion and attention mechanisms, enhancing the model's perceptual capabilities and its ability to extract key information.

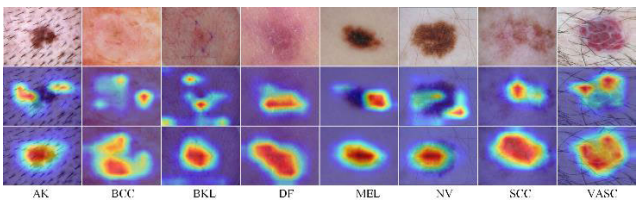


FIGURE 9. Visualization comparison of classification results of ConvNeXt (second row) and the proposed EFAM-Net model (third row), based on ISIC 2019 images (first row).

V. CONCLUSION

This paper has proposed a novel classification model, named EFAM-Net, utilizing a newly designed Attention Residual Learning ConvNeXt (ARLC) block, a Parallel ConvNeXt (PCNXt) block, and a Multi-scale Efficient Attention Feature Fusion (MEAFF) block to extract complex feature information from skin lesion images. The ARLC block is used in the shallow network of the model, as a replacement of the original ConvNeXt block, in order to enhance the model's ability to extract low-level features, such as colors and textures, and improve its anti-noise ability for successful extraction of effective feature information from images with noise such as artifacts. The ConvNeXt block in the deep network of the model is replaced by the PCNXt block. The parallel branch structure of the PCNXt block provides global information for the model, which effectively helps improve its ability of locating the lesion areas. The MEAFF block allows the model to successfully focus on the difference of the lesion areas. Overall, the ARLC block provides the model with general features such as image location information, the PCNXt block helps the model extract more complex features with almost no increase in its complexity, and the MEAFF block allows to achieve feature reuse, which further improves the classification performance of the proposed model.

For visualization of the classification results of the proposed EFAM-Net model, the Gradient Weighted Class Activation Mapping (Grad-CAM) technique has been used to generate attention visualization maps and analyze the classification results. In addition, multiple sets of experiments were conducted using three different datasets, including ablation study experiments, and classification performance comparison experiments with baseline models and state-of-the-art models. The ablation study results have shown that all newly designed blocks had a positive effect on the classification performance of the proposed EFAM-Net model, reaching the highest metric values when all these blocks were integrated together into the model. The classification performance comparison experiments with other existing models have shown that on the most representative ISIC 2019 dataset, the overall accuracy, macro-averaged accuracy, and the average recall, specificity, and F1 score of the proposed EFAM-Net model achieved top results among them.

Nonetheless, there are still some unsolved problems in the tasks of skin lesion classification. For example, models still need effective dataset preprocessing and data augmentation to demonstrate better results. How to realize an efficient use of datasets is worth further study. In the future, we will extend the proposed model evaluation to more diverse datasets and datasets with different imaging modalities, which can further verify the generalizability of EFAM-Net for various skin conditions.

ETHICAL STATEMENT

Ethical review and approval were waived for this study as the utilized data only involved anonymous pictures of skin lesions with informed consent given by all subjects involved. In addition, no animals were used in conducting the experiments.

REFERENCES

[1] S. Kumari, P. K. Choudhary, R. Shukla, A. Sahebkar, and P. Kesharwani, "Recent advances in nanotechnology based combination drug therapy for skin cancer," *J. Biomater. Sci., Polym. Ed.*, vol. 33, no. 11, pp. 1435–1468, Jul. 2022.

[2] L. Zeng, B. H. J. Gowda, M. G. Ahmed, M. A. S. Abourehab, Z.-S. Chen, C. Zhang, J. Li, and P. Kesharwani, "Advancements in nanoparticle-based treatment approaches for skin cancer therapy," *Mol. Cancer*, vol. 22, no. 1, p. 10, Jan. 2023.

[3] M. E. Celebi, N. Codella, and A. Halpern, "Dermoscopy image analysis: Overview and future directions," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 2, pp. 474–478, Mar. 2019.

[4] J. Lin, S. Han, L. Cui, Z. Song, M. Gao, G. Yang, Y. Fu, and X. Liu, "Evaluation of dermoscopic algorithm for seborrhoeic keratosis: A prospective study in 412 patients," *J. Eur. Acad. Dermatol. Venereol.*, vol. 28, no. 7, pp. 957–962, Jul. 2014.

[5] L. J. Loescher, M. Janda, H. P. Soyer, K. Shea, and C. Curiel-Lewandrowski, "Advances in skin cancer early detection and diagnosis," *Seminars Oncol. Nursing*, vol. 29, no. 3, pp. 170–181, Aug. 2013.

[6] S. Qasim Gilani, T. Syed, M. Umair, and O. Marques, "Skin cancer classification using deep spiking neural network," *J. Digit. Imag.*, vol. 36, no. 3, pp. 1137–1147, Jan. 2023.

[7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015, *arXiv:1409.1556*.

[8] J. P. Villa-Pulgarin, A. A. Ruales-Torres, D. Arias-Garzón, M. A. Bravo-Ortiz, H. B. Arteaga-Arteaga, A. Mora-Rubio, J. A. Alzate-Grisales, E. Mercado-Ruiz, M. Hassaballah, S. Orozco-Arias, O. Cardona-Morales, and R. Tabares-Soto, "Optimized convolutional neural network models for skin lesion classification," *Comput., Mater. Continua*, vol. 70, no. 2, pp. 2131–2148, 2022.

[9] 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. 2261–2269.

[10] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. AAAI Conf. Artif. Intell.*, 2017, vol. 31, no. 1, pp. 4278–4284.

[11] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[12] X. Zeng, Z. Ji, H. Zhang, R. Chen, Q. Liao, J. Wang, T. Lyu, and L. Zhao, "DSP-KD: Dual-stage progressive knowledge distillation for skin disease classification," *Bioengineering*, vol. 11, no. 1, p. 70, Jan. 2024.

[13] R. Maurya, A. K. Bais, T. Gopalakrishnan, M. K. Dutta, N. N. Pandey, and Y. V. S. Murthy, "Skin lesion classification using deep feature fusion and selection using XGBoost classifier," in *Proc. IEEE Int. Students' Conf. Electr., Electron. Comput. Sci. (SCEECS)*, vol. 64, Feb. 2024, pp. 1–5.

[14] A. Mahbod, G. Schaefer, C. Wang, G. Dorffner, R. Ecker, and I. Ellinger, "Transfer learning using a multi-scale and multi-network ensemble for skin lesion classification," *Comput. Methods Programs Biomed.*, vol. 193, Sep. 2020, Art. no. 105475.

[15] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.

[16] 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.

[17] 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.

[18] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1800–1807.

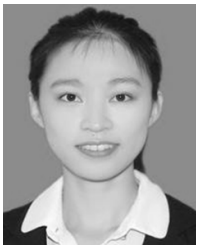
[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[20] Z. Lan, S. Cai, X. He, and X. Wen, "FixCaps: An improved capsules network for diagnosis of skin cancer," *IEEE Access*, vol. 10, pp. 76261–76267, 2022.
[21] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.
[22] R. Karthik, T. S. Vaichole, S. K. Kulkarni, O. Yadav, and F. Khan, "Eff2Net: An efficient channel attention-based convolutional neural network for skin disease classification," *Biomed. Signal Process. Control*, vol. 73, Mar. 2022, Art. no. 103406.
[23] M. Tan and Q. Le, "EfficientNetV2: Smaller models and faster training," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 10096–10106.
[24] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11531–11539.
[25] X. Liu, L. Yang, X. Ma, and H. Kuang, "Skin disease classification based on multi-level feature fusion and attention mechanism," in *Proc. IEEE 3rd Int. Conf. Inf. Technol., Big Data Artif. Intell. (ICIBA)*, vol. 3, May 2023, pp. 1500–1504.
[26] 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.
[27] J. Zhang, Y. Xie, Y. Xia, and C. Shen, "Attention residual learning for skin lesion classification," *IEEE Trans. Med. Imag.*, vol. 38, no. 9, pp. 2092–2103, Sep. 2019.
[28] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Training very deep networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, pp. 2377–2385, 2015.
[29] M. Combalia, N. C. F. Codella, V. Rotemberg, B. Helba, V. Vilaplana, O. Reiter, C. Carrera, A. Barreiro, A. C. Halpern, S. Puig, and J. Malvehy, "BCN20000: Dermoscopic lesions in the wild," 2019, *arXiv:1908.02288*.
[30] N. C. F. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 168–172.
[31] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Sci. Data*, vol. 5, no. 1, pp. 1–9, Aug. 2018.
[32] I. Loshchilov, "Decoupled weight decay regularization," 2017, *arXiv:1711.05101*.
[33] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.
[34] S. Visa, B. Ramsay, A. L. Ralescu, and E. Van Der Knaap, "Confusion matrix-based feature selection," *Maics*, vol. 710, no. 1, pp. 120–127, 2011.
[35] A. H. Jui, S. Sharnami, and A. Islam, "A CNN based approach to classify skin cancers using transfer learning," in *Proc. 25th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2022, pp. 1063–1068.
[36] M. A. Khan, M. Sharif, T. Akram, S. Kadry, and C. Hsu, "A two-stream deep neural network-based intelligent system for complex skin cancer types classification," *Int. J. Intell. Syst.*, vol. 37, no. 12, pp. 10621–10649, Dec. 2022.
[37] S. Benyahia, B. Meftah, and O. Lézoray, "Multi-features extraction based on deep learning for skin lesion classification," *Tissue Cell*, vol. 74, Feb. 2022, Art. no. 101701.
[38] Y. Wang, Y. Wang, J. Cai, T. K. Lee, C. Miao, and Z. J. Wang, "SSD-KD: A self-supervised diverse knowledge distillation method for lightweight skin lesion classification using dermoscopic images," *Med. Image Anal.*, vol. 84, Feb. 2023, Art. no. 102693.
[39] S. Gopikha and M. Balamurugan, "Regularised layerwise weight norm based skin lesion features extraction and classification," *Comput. Syst. Sci. Eng.*, vol. 44, no. 3, pp. 2727–2742, 2023.
[40] M. Sharafudeen and S. S. V. Chandra, "Detecting skin lesions fusing handcrafted features in image network ensembles," *Multimedia Tools Appl.*, vol. 82, no. 2, pp. 3155–3175, Jan. 2023.
[41] I. Pacal, M. Alaftekin, and F. D. Zengul, "Enhancing skin cancer diagnosis using Swin transformer with hybrid shifted window-based multi-head self-attention and SwiGLU-based MLP," *J. Imag. Informat. Med.*, pp. 1–19, Jun. 2024.
[42] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.
[43] T. M. Alam, K. Shaukat, W. A. Khan, I. A. Hameed, L. A. Almuqren, M. A. Raza, M. Aslam, and S. Luo, "An efficient deep learning-based skin cancer classifier for an imbalanced dataset," *Diagnostics*, vol. 12, no. 9, p. 2115, Aug. 2022.
[44] L. Hoang, S.-H. Lee, E.-J. Lee, and K.-R. Kwon, "Multiclass skin lesion classification using a novel lightweight deep learning framework for smart healthcare," *Appl. Sci.*, vol. 12, no. 5, p. 2677, Mar. 2022.
[45] 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.
[46] H. Wang, Q. Qi, W. Sun, X. Li, B. Dong, and C. Yao, "Classification of skin lesions with generative adversarial networks and improved MobileNetV2," *Int. J. Imag. Syst. Technol.*, vol. 33, no. 5, pp. 1561–1576, Sep. 2023.
[47] L. Liu, X. Zhang, and Z. Xu, "An adaptive weight search method based on the grey wolf optimizer algorithm for skin lesion ensemble classification," *Int. J. Imag. Syst. Technol.*, vol. 34, no. 2, Mar. 2024, Art. no. e23049.
[48] G. P. Devaraj and R. Ravi, "Advancing skin cancer diagnosis with a multi-branch ShuffleNet architecture," *Int. J. Imag. Syst. Technol.*, vol. 34, no. 2, Mar. 2024, Art. no. e23051.
[49] A. K. Gairola, V. Kumar, A. K. Sahoo, M. Diwakar, P. Singh, and D. Garg, "Multi-feature fusion deep network for skin disease diagnosis," *Multimedia Tools Appl.*, pp. 1–26, Mar. 2024.

ZHANLIN JI (Member, IEEE) received the Ph.D. degree from the University of Limerick, Ireland, in 2010. Currently, he is a Professor with Zhejiang Agriculture and Forestry University, China, and an Associate Researcher with the Telecommunications Research Centre (TRC), University of Limerick. He has authored/co-authored more than 100 research papers in refereed journals and conferences. His research interests include ubiquitous consumer wireless world (UCWW), the Internet of Things (IoT), cloud computing, big data management, and artificial intelligence (AI)-based image processing. He was a recipient of Irish Research Council for Science, Engineering, and Technology (IRCSET) Post-Graduate Research Scholarship, in 2008, and the IRC Postdoctoral Fellowship, in 2013.

XUAN WANG was born in 1999. He received the B.S. degree from Tangshan University, in 2022. He is currently pursuing the master's degree with North China University of Science and Technology. His research interests include machine vision and graphic image processing.

CHUNLING LIU received the Ph.D. degree from the Peking Union Medical College, Chinese Academy of Medical Sciences, in 2014, specializing in cell research. She is currently an Attending Physician with the Pathology Department, Tangshan People's Hospital.

ZHIWU WANG received the Ph.D. degree from Tianjin Medical University, in 2014. He is currently a Chief Physician with the Second Department of Radiotherapy and Chemotherapy, Tangshan People's Hospital. He is engaged in the comprehensive medical treatment of lung cancer and digestive system tumors. His current research interests include medical data processing and screening tumor immunotherapy effect prediction markers based on artificial intelligence.



NA YUAN received the bachelor's degree from Heilongjiang University, in July 2011, and the master's degree from Hebei University of Technology, in January 2014. She is currently a Lecturer with Tangshan University. Her research interests include intelligent control, machine vision, and graphic image processing.

IVAN GANCHEV (Senior Member, IEEE) received the Engineering and Ph.D. degrees (summa cum laude) from Saint Petersburg University of Telecommunications, in 1989 and 1995, respectively. He is an International Telecommunications Union (ITU-T) Invited Expert and an Institution of Engineering and Technology (IET) Invited Lecturer, currently affiliated with the University of Limerick, Ireland; the University of Plovdiv "Paisii Hilendarski," Bulgaria; and the Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Bulgaria. He was involved in more than 40 international and national research projects and has served on the TPC of more than 400 prestigious international conferences/symposia/workshops. He has authored/co-authored one monographic book, three textbooks, four edited books, and more than 300 research papers in refereed international journals, books, and conference proceedings. He is on the editorial board of and has served as a guest editor for multiple prestigious international journals.