Signal, Image and Video Processing (2024) 18:6353–6361
https://doi.org/10.1007/s11760-024-03321-y

ORIGINAL PAPER

Skin lesion classification from dermoscopy images using ensemble learning of ConvNeXt models

Elif Baykal Kablan[1] · Selen Ayas[2]

Received: 29 May 2023 / Revised: 9 February 2024 / Accepted: 25 May 2024 / Published online: 6 June 2024

Abstract

Automatic classification of dermoscopy images plays a crucial role in the early diagnosis and treatment of serious diseases like skin cancer. However, it poses several challenges, including similar appearance lesions, different types of skin structures, variations in lesion stages, insufficient or inaccurate data, and artifacts present in dermoscopy images. In skin lesion classification tasks, deep learning-based methods have recently demonstrated superior performance compared to traditional machine learning-based methods. In this study, a novel ensemble-based approach is designed for skin lesion classification by leveraging the diverse information captured by different architectures of ConvNeXt models which have been demonstrated to achieve comparable performance to most vision transformers by utilizing a CNN backbone. More specifically, firstly, different versions of pre-trained and fine-tuned ConvNeXt models, namely Tiny, Small, Base, and Large, were used for the classification of skin lesion images to analyze and compare classification performances on the publicly available ISIC 2019 dataset. Among the individual models, ConvNeXt-Large achieved the highest accuracy rate of 97.2%, making it the top-performing model. Then, all four ConvNeXt models were fused using confidence scores to improve classification accuracy. The ensemble approach achieved an overall classification accuracy of 97.7%, surpassing both the performance of individual models and state-of-the-art methods. Additionally, a sensitivity value of 84.2% and a specificity value of 97.9% were obtained. The findings of this study provide evidence that the proposed approach effectively and accurately classifies skin lesions from dermoscopy images.

Keywords Skin lesion classification · Skin cancer · Dermoscopy image analysis · ConvNeXt · Ensemble learning

1 Introduction

Skin cancer is one of the most common types of cancer that occurs due to abnormal growths in the skin cells [1]. According to the World Health Organization (WHO), skin cancer affects roughly three million individuals globally every year, resulting in thousands of deaths [2]. Regular skin examinations by expert dermatologists and awareness of changes in nevus or skin spots are essential for early diagnosis of potential skin cancer. This allows for the treatment of cancer cells before they spread to surrounding tissues. Additionally, when skin cancer is diagnosed early, the treatment process is both easier and less invasive [3].

Dermoscopy is a diagnostic technique that does not involve any invasive procedures and enables the visualization of skin lesions with higher magnification and improved clarity [4]. It is commonly used by dermatologists in the diagnosis of melanoma and other skin cancer types. However, this technique is known to be time-consuming, tiring, and prone to errors and variations in diagnosis among dermatologists [5]. Therefore, there is a demand for computer-aided diagnosis (CAD) systems to reduce diagnostic subjectivity and improve accuracy and consistency. These systems can also aid in the early detection and treatment of skin cancer by identifying early-stage skin lesions that may be missed by the naked eye [6].

The purpose of automated CAD systems is to categorize skin lesions as malignant or benign, sometimes even precisely categorizing these two classes into their own sub-

✉ Elif Baykal Kablan
ebaykal@ktu.edu.tr

Selen Ayas
selenayas@ktu.edu.tr

[1] Department of Software Engineering, Karadeniz Technical University, 61080 Trabzon, Turkey

[2] Department of Computer Engineering, Karadeniz Technical University, 61080 Trabzon, Turkey

classes. On the other hand, the classification of skin lesions poses several challenges that can result in misdiagnosis. Some of these challenges include:

1. Similar-looking lesions: Some skin lesions from different classes may have a similar appearance.
2. Differences in skin characteristics: Different individuals may have different skin types and structures, causing skin lesions from the same class to look different.
3. Variations in lesion stages: An early-stage lesion may have a different appearance from a later-stage lesion.
4. Insufficient or inaccurate data: The data used for skin lesion classification may be insufficient (e.g., for a rare type of lesion) or inaccurate.
5. Artifacts: Artifacts, including hair, skin lines, and blood vessels, may be present in dermoscopy images.

With the rapid progress of deep learning technology, it has become the preferred method for medical image analysis in computer vision [7, 8]. Compared to traditional classification methods, deep learning has exhibited enhanced robustness and superior generalization capability. One of the most well-known deep learning models, Convolutional Neural Networks (CNNs) [9] are excellent at capturing spatial information and detecting local patterns, making them suitable for image analysis tasks, including skin lesion classification. However, as higher performance and scalability demand increased, researchers explored new architectures such as Vision Transformers (ViTs) [10]. ViTs introduced the concept of self-attention, allowing models to capture global dependencies in input images. This contribution led to remarkable improvements in image classification tasks by dealing with the complexities of aforementioned challenges of the image datasets. In 2022, Liu et al. [11] introduced the ConvNeXt model, which combines the strengths of CNNs and Transformers. ConvNeXt utilizes a CNN backbone to capture local features and an attention mechanism to capture global dependencies. This architecture has been shown to surpass the performance of traditional transformers and even the successful ViT model, Swin Transformer [12], while overcoming the limitations of input size.

Ensemble methods have gained significant popularity in diverse medical image classification tasks. [13, 14]. The classifiers with different architectures used in ensemble methods can capture image information at different levels, leading to more accurate decisions. To our knowledge, there is no existing study on the classification of skin lesions from dermoscopy images using ConvNeXt models. On the other hand, this study is the first to utilize both individual ConvNeXt models and ensemble learning technique for classifying skin lesions from dermoscopy images. Therefore, the main contributions of this study are as follows:

1. This is the pioneering study that applies ConvNeXt model architectures to dermoscopy images for the task of skin lesion classification.
2. We conducted experiments without altering the existing structures of ConvNeXt models (Tiny, Small, Base, Large) to enable effective transfer learning for eight-class skin lesion classification.
3. We investigated the effect of ensemble learning, and the results demonstrated that the ensemble of different ConvNeXt models outperformed individual models in the classification tasks.
4. For both individual models and ensemble models, five-fold cross-validation and testing were performed to evaluate their performance. The ensemble of all ConvNeXt models achieved an overall classification accuracy of 97.7%, surpassing both the performance of individual models and state-of-the-art methods.
5. To ensure the validity of this study, comparisons were made with state-of-the-art methods based on CNNs [15–19] and Vision Transformer (ViT) models [20]. These methods were selected as they represent the most frequently compared approaches in the recent literature. Training and testing processes were conducted on the publicly available ISIC 2019 dataset, commonly used for skin lesion classification. This allowed for a fair comparison of the proposed approach against other state-of-the-art methods.

Based on our findings, this study highlights the potential of ConvNeXt models in accurately classifying skin lesions from dermoscopy images. Further research in this direction can contribute to the development of more effective and reliable automated systems for skin lesion analysis.

2 Related work

The initial studies on skin lesion classification in the literature considered the lesion classification problem as a binary classification problem, where the lesions were categorized as either malignant or benign. With the emergence of larger datasets [21–23] that include subtypes of malignant and benign lesions, recent studies have focused more on automated multi-class skin lesion classification. However, automated multi-class classification of skin lesions remains a challenging task due to the challenges mentioned in Sect. 1 and the existence of multiple classes.

Deep Learning has garnered significant attention in the field of medical image classification, including the classification of skin lesions. Extensive research has been conducted, employing numerous deep learning approaches to tackle this task. Esteva et al. [24] utilized the GoogleNet Inception v3 model to train on a dataset consisting of 129,450 clinical images, encompassing 2,032 different diseases. The proposed model achieved performance comparable to that of all tested experts and demonstrated the ability of artificial intelligence to classify skin cancer at a level similar to dermatologists. Abbas and Celebi [25] proposed a new classification method named DermoDeep, which combines various visual features and deep neural network approaches to classify pigmented skin lesions. They evaluated the method on 2800 region-of-interests (ROIs) and achieved an AUC of 0.96, with a sensitivity of 93% and specificity of 95%. Gessert et al. [17] proposed an ensemble of deep learning models comprising EfficientNets, SENet, and ResNeXt WSL, which were selected using a search strategy. They addressed the class imbalance issue with a loss balancing approach. The results showed that EfficientNets models performed well on the ISIC2019 dataset. Furthermore, the automatic selection of the ensemble of SENet154 and ResNext models indicated that the variability in network architectures yielded better results. Pacheco and Krohling [26] highlighted the potential for achieving improved performance by considering the demographic characteristics of the patient, rather than solely relying on the classification of skin lesions based on images. To this end, they proposed a new approach called MetaBlock, which uses the most relevant features and metadata. The results showed that the MetaBlock approach improved classification for all tested models. Kassem et al. [18] tested a modified GoogleNet model using transfer learning approach on the ISIC2019 dataset. The proposed model achieved the following classification metrics: accuracy of 94.92%, sensitivity of 79.8%, specificity of 97%, and precision of 80.36%. Molina-Molina et al. [15] presented an approach that combines deep learning features extracted from Densenet-201 with 1D fractal signatures of texture-based features through transfer learning. The proposed method achieved an accuracy of 97.35%, sensitivity of 66.45%, and specificity of 97.85% on the ISIC2019 dataset. Iqbal et al. [19] proposed a Deep Convolutional Neural Network (DCNN) model with fewer filters and parameters to improve efficacy and performance. The proposed model achieved an accuracy of 89.58%, sensitivity of 89.58%, and specificity of 97.57% on the ISIC2019 dataset. Zhao et al. [16] presented a new skin lesion image classification approach based on SLA-StyleGAN, a specific image augmentation method for skin lesions, using the DenseNet201 architecture. Additionally, they introduced a novel loss function that aims to increase the distance between samples from different classes while reducing the distance between samples within the same class. Experimental results demonstrated that the proposed framework achieved a balanced multi-class accuracy of 93.64% on the ISIC2019 dataset. Ayas [20] proposed the first vision transformer-based model for multi-class skin lesion image classification. The proposed Swin Transformer model achieved a sensitivity of 82.3%, specificity of 97.9%, accuracy of 97.2%, and balanced accuracy of 82.3% on the ISIC2019 dataset.

In this paper, we presented the effectiveness of ConvNeXt [11] model, which combines the strengths of CNNs and Transformers, in skin lesion classification. The ConvNeXt is a CNN-based model and it has been proposed to improve the performance of vision transformers. Unlike vision transformers, ConvNext does not rely on specialized modules such as shifting window attention or relative position biases, resulting in a more modern model that achieves comparable performance, memory usage, and FLOPs (floating-point operations per second) to the Swin Transformer [27]. To the best of the author's knowledge, this is the first study to utilize the ConvNext model for multi-class skin lesion classification. The experimental results demonstrated that the proposed approach achieved better performance for both individual and ensemble models in terms of sensitivity, specificity, and accuracy metrics.

3 Methods

3.1 ConvNeXt

The ConvNeXt architecture [11], proposed by Liu et al. in 2022, aims to outperform the performance of ViTs. To achieve this goal, it takes advantage of attention-based classifiers and conventional ResNet model. Motivated by the need to capture global dependencies and contextual information, the ConvNeXt architecture employs convolutions with large receptive fields as its fundamental building block. Additionally, as a pure CNN architecture, ConvNeXt outperforms Swin Transformer, the most powerful transformer model on the ImageNet-1K dataset [28]. The ConvNeXt architecture is shown in Fig. 1.

ConvNeXt has a structure that is very similar to ResNet50, consisting of a head feature extraction layer, a middle layer characterized by a bottleneck structure encompassing four different dimensions, and a high-dimensional feature classification layer. However, the interior of each layer and the strategy of stacking have undergone several changes. First, the stacking number of each block has been revised from 3:4:6:3 to 3:3:9:3, which is similar to the transformer model. Within each ConvNeXt block, there is a depth-wise convolution operation, which is then accompanied by 1×1 convolutions. To achieve this, the depth-wise convolution adopts a group-wise convolution approach that involves grouping the channels together. Secondly, the bottleneck design has been modified to following sequence of operations: firstly, it performs feature extraction, followed by dimension reduction, and finally, dimension expansion. Thirdly, the size of the convolution kernel has been changed from 3x3 to 7x7. Fourthly, the activation function has been replaced from Rectified Lin-

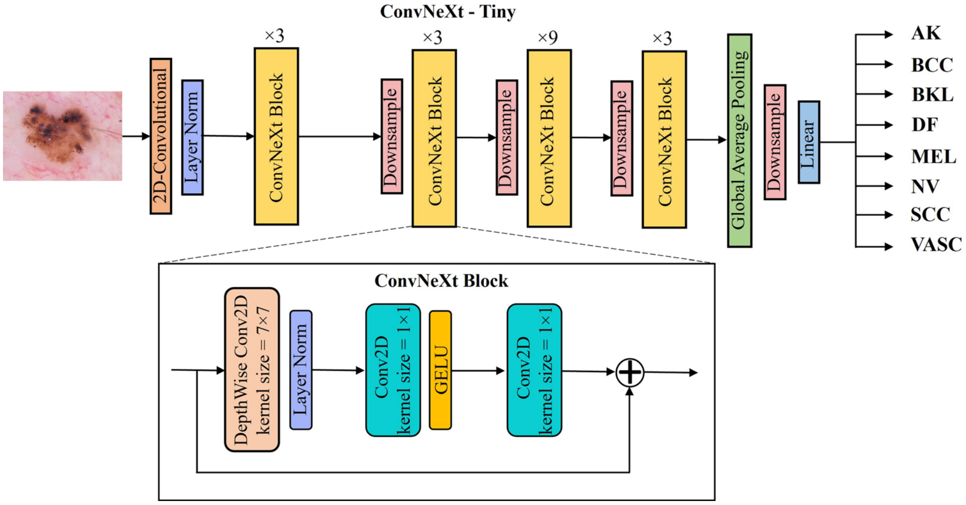

Fig. 1 The architecture of the ConvNeXt-Tiny model. The downsample layer and ConvNeXt block are stacked in the ratio of 3:3:9:3 ratio of 4 stages. The GELU represents the Gaussian Error Linear Unit. The output class names are abbreviated as AK: actinic keratosis, BCC: basal cell carcinoma, BKL: benign keratosis, DF: dermatofibroma, NV: melanocytic nevus, MEL: melanoma, SCC: squamous cell carcinoma, and VASC: vascular lesion, respectively

Table 1 The configurations of the four ConvNeXt model versions

Model	Number of channels	Number of blocks
Tiny (T)	(96, 192, 384, 768)	(3, 3, 9, 3)
Small (S)	(96, 192, 384, 768)	(3, 3, 27, 3)
Base (B)	(128, 256, 512, 1024)	(3, 3, 27, 3)
Large (L)	(192, 384, 768, 1536)	(3, 3, 27, 3)

ear Unit (ReLU) to Gaussian Error Linear Unit (GELU), with fewer activation functions used. Finally, a notable change is the adoption of layer normalization instead of batch normalization as well as employing fewer normalization layer. These modifications, along with new parameters, structures, and functions, have gradually improved the performance of ConvNeXt, even outperforming the ViT such as Swin Transformer.

Additionally, four versions of ConvNeXt are proposed, namely, ConvNeXt-Tiny (T), ConvNeXt-Small (S), ConvNeXt-Base (B), and ConvNeXt-Large (L). The diversity of these versions varies as the number of channels and blocks used in each stage differs, as shown in Table 1.

3.2 The proposed ensemble of ConvNeXt classifiers

Ensemble learning is a powerful technique widely used in computer vision, where different classifiers are combined to enhance classification performance. By leveraging the diverse information captured by classifiers with different architectures, ensemble models have the potential to achieve higher accuracy compared to individual base learners. This approach is commonly employed in various medical image classification tasks [13, 14]. In this study, all versions of the ConvNeXt model (ConvNeXt-T, ConvNeXt-S, ConvNeXt-B, ConvNeXt-L) are selected as the base classifiers of the ensemble model.

Let $x(w, h, c)$ be an unseen test image with a size of $w \times h$ pixels and c channels. To classify x, we utilize the following approach. In the final decision step, each individual fine-tuned ConvNeXt classifier C_i in the ensemble C produces confidence scores of the input x belonging to the class y membership as given in (1). We then select the class with high confidence value as the label for x as given in (2).

$$P_y(x) = \sum_{C_i \epsilon C} P_{C_i, y}(x) \tag{1}$$

$$C(x) = arg\, \max_{y \epsilon Y} P_y(x) \tag{2}$$

4 Experimental setup and results

All experiments were conducted on a computer equipped with an Intel(R) Core(TM) i9-11900K 3.50 GHz CPU and an NVIDIA GeForce RTX 3080 12GB GPU. The ConvNeXt models were developed using the PyTorch deep learning library.

4.1 ISIC2019 Skin lesion classification dataset

The ISIC 2019 skin lesion dataset [21–23, 29, 30] is a dermatology dataset created by the International Skin Imaging Collaboration (ISIC) in 2019. It is specifically designed for skin cancer diagnosis and consists of a total of 25,331 images belonging to 8 subcategories of both benign and malignant skin lesions. The subcategories are named as follows: actinic keratosis (AK), basal cell carcinoma (BCC), benign keratosis (BKL), dermatofibroma (DF), melanocytic nevus (NV), melanoma (MEL), squamous cell carcinoma (SCC), and vascular lesion (VASC). Figure 2 includes some sample images of the dataset in each lesion category. The dataset does not provide ground truth labels for the test data. To make a fair

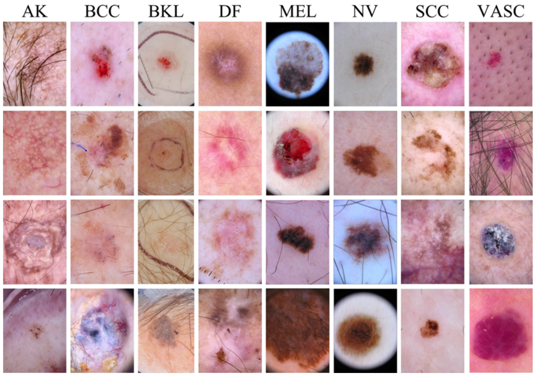

Fig. 2 Some sample images of the dataset in each leasion category. AK: actinic keratosis, BCC: basal cell carcinoma, BKL: benign keratosis, DF: dermatofibroma, NV: melanocytic nevus, MEL: melanoma, SCC: squamous cell carcinoma, VASC: vascular lesion

comparison with state-of-the-art methods we followed the same training/testing protocol presented in [20]. We divided the available training data into three subsets: training, validation, and test, with a split ratio of 70%, 10%, and 20%, respectively. We also applied the 5-fold cross-validation technique, where the dataset was split into 5 folds, keeping the number of images of the same class in each fold equal. This ensures to avoid problems such as all samples being from one class or certain classes not being represented. Table 2 presents the number of training, validation, and test samples in each lesion category.

We employed data augmentation techniques during the training process to enhance the model's generalization ability. The augmentation techniques include a range of transformations, including geometric transformations like random horizontal and vertical flips, random rotation, and color jitter transformations that involve brightness, contrast, and saturation adjustments. Additionally, we resized all the images to 224×224 pixels to ensure consistency in input dimensions during training.

4.2 Training details

The ISIC 2019 dataset exhibits class imbalance, with the NV class containing over 12,000 images whereas classes such as AK, DF, SVC, and VASC comprise a smaller number of images ranging from 200 to 900. Imbalanced datasets tend to bias the model towards the class with a larger number of samples, which can result in an increase in false positives (FP) or false negatives (FN) depending on the imbalance. In this study, to address the issue of overfitting on the NV class during training, a weighting scheme based on inverse class frequency is applied to the cross-entropy loss function. The weight value for each class, $weight_{C_i}$, is calculated using the Eq. 3.

Table 2 The number of training, validation, and test samples in each lesion category

	AK	BCC	BKL	DF	MEL	NV	SCC	VASC
Training	607	2326	1837	168	3165	9012	440	177
Validation	87	333	263	24	453	1288	63	26
Test	173	664	524	47	904	2575	125	50

AK actinic keratosis, *BCC* basal cell carcinoma, *BKL* benign keratosis, *DF* dermatofibroma, *NV* melanocytic nevus, *MEL* melanoma, *SCC* squamous cell carcinoma, *VASC* vascular lesion

$$weight_{C_i} = \frac{\sum_{j=1}^{k} N_j}{k \times N_i} \tag{3}$$

where N_i denotes the number of images in ith class and k denotes the number of class.

The training data in the ISIC 2019 dataset is not sufficient for training CNN-based architectures from scratch. Therefore, in this study, instead of training ConvNeXt models from scratch, pre-trained models on the 1K-class ImageNet dataset were fine-tuned as skin lesion classifiers. During the training of the models, the AdamW optimization method was applied with a learning rate of 1e-5 and a weight decay value of 1e-8. The cross-entropy loss function was used as the error function. The batch size was set to 8, and the number of epochs was set to 50 for all individual and ensemble models.

4.3 Performance metrics

The classification performance of the models was evaluated considering three widely used quantitative metrics, i.e., Sensitivity, Specificity, and Accuracy. The study was considered as a multi-class (c) classification problem, where each test sample is assigned to one of the predefined classes $Class_1$, $Class_2$,..., $Class_c$. The confusion matrix [31] is used to analyse the results of the multi-class classifier. The confusion matrix shows the relationship between the actual class values and the class values predicted by the classifier. The confusion matrix for the c-class problem can be expressed as a $c \times c$ table where each cell $x_{i,j}$, $(i = 1, ..., c$ and $j = 1, ..., c)$ of the confusion matrix at row i and column j provides the number of instances for which the predicted class is j and the actual class is i. A binary confusion matrix is a special case when there are only two classes. Hence, a $c \times c$ confusion matrix can be represented as a set of c binary confusion matrices, one for each class. Table 3 represents a confusion matrix for a c-class problem.

Sensitivity measures the ability of a classification model to correctly identify positive instances out of all actual positive instances whereas specificity measures the ability of a classification model to correctly identify negative instances out of all actual negative instances in a dataset. Accuracy measures the overall correctness of a classification model across

Table 3 Confusion matrix used to calculate evaluation metrics

		Classifier output				
		$Class_1$	$Class_2$	$Class_3$	$\cdots$	$Class_c$
Ground truth	$Class_1$	x_{11}	x_{12}	x_{13}	$\cdots$	x_{1c}
	$Class_2$	x_{21}	x_{22}	x_{23}	$\cdots$	x_{2c}
	$Class_3$	x_{31}	x_{32}	x_{33}	$\cdots$	x_{3c}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	$Class_c$	x_{c1}	x_{c2}	x_{c3}	$\cdots$	x_{cc}

all classes. The sensitivity, specificity, and accuracy metrics for class$_i$ are formulated as follows:

$$Sensitivity_{\text{class}_i} = \frac{x_{ii}}{x_{ii} + \sum_{j=1}^{c} x_{ij}} \tag{4}$$

$$Specificity_{\text{class}_i} = \frac{\sum_{j \neq i}^{c} \sum_{k \neq i}^{c} x_{jk}}{\sum_{j \neq i}^{c} \sum_{k \neq i}^{c} x_{jk} + \sum_{j \neq i}^{c} x_{ij}} \tag{5}$$

$$Accuracy_{\text{class}_i} = \frac{x_{ii} + \sum_{j \neq i}^{c} \sum_{k \neq i}^{c} x_{jk}}{\sum_{i=1}^{c} \sum_{j=1}^{c} x_{ij}} \tag{6}$$

Accuracy, sensitivity, and specificity measures are calculated separately for each class using the confusion matrix obtained. In the study, the class for which the classification performance is to be calculated was defined as the positive class, while all other classes were defined as negative classes. As a result, the overall classification performance measure was obtained by averaging the c classes.

4.4 Results

First, the individual and ensemble performances of four different versions of the ConvNeXt model, as stated in Table 1, were analyzed for skin lesion classification. The mean and standard deviation results of each model obtained through 5-fold cross-validation are provided in Table 4. The classification performance was evaluated using the three metrics described in Sect. 4.3: accuracy, sensitivity, and specificity. The results show that by increasing the model complexity in the order of tiny, small, base, and large, the SE metric improves from 80.5% to 81.3%. Furthermore, it has been observed that the transfer learning approach achieves an accuracy of over 96% for all ConvNeXt models. Additionally, ensemble models have increased the highest sensitivity value obtained in individual models from 81.3 to 84.2. The proposed ConvNeXt T-S-B-L (overall) ensemble method achieved the best values with accuracy of 97.7%, sensitivity of 84.2%, and specificity of 97.9%. Furthermore, the high average results and low standard deviation values indicate that the models generally perform well and the results are consistent. In conclusion, such results have demonstrated the effectiveness of both individual and ensemble architecture of ConvNeXt models.

We also compared the effectiveness and robustness of the ConvNeXt models with state-of-art methods: Molina's method [15], Zhao's method [16], EfficientNets [17], Kassem's method [18], CSLNet [19], and Swin transformer-based models [20]. We chose these methods because they are the most frequently compared studies in the literature. For a fair comparison, we used the same configuration of the dataset as in [20]. We also applied 5-fold cross-validation to avoid the variability of samples which may affect the performance of the models. Quantitatively, Table 5 summarizes the classification performance of the proposed method with six state-of-the-art methods on the ISIC 2019 dataset. The symbol "-" refers unreported results. The highest accuracy of 97.7% was achieved with the ensemble of all ConvNeXt models. Additionally, it can be observed that the sensitivity and specificity values are also high. Molina et al. [15] achieved an average sensitivity value of 66.5% by using the entire dataset without performing a specific training-test split. However, they reported that the low sensitivity values for classes like DF, SCC, and VASC were attributed to the limited number of images available for these classes. Zhao et al. [16] were able to increase the sensitivity value to 68.2% by incorporating various contributions. Gessert et al. [17] achieved the best sensitivity value of 72.5% by using an additional dataset. Kassem et al. [18] addressed the imbalanced dataset problem and achieved a sensitivity of 79.8% by using only 191 images. Iqbal et al. [19], addressed the issue of class imbalance in the dataset and achieved an impressive sensitivity using their proposed CSLNet model. However, the classification accuracy of the model was considerably low. Ayas et al. [20] obtained state-of-the-art classification results by using different sub-versions of the Swin transformers. Table 5 further demonstrates that the proposed ConvNext models in the study yield competitive results against the Swin Transformer models. These findings highlight that different models may be effective in different scenarios. Researchers can contribute to a better understanding of the strengths of each approach and guide future studies by conducting in-depth analysis of these competitive results. This competition encourages progress and fosters continuous innovation to achieve better performance.

Figure 3 shows the confusion matrices obtained for individual and ensemble models on the fold-1 test set. The diagonal values in the confusion matrices represent the ratio of correctly classified samples to the total number of samples in each class, giving the sensitivity values for each class. As can be seen from Fig. 3, the ConvNeXt-T model achieved 68% sensitivity for the AK class in the fold-1 test set, which was increased to 78% with the ensemble model. Similarly, the classification of MEL class images obtained 74% accu-

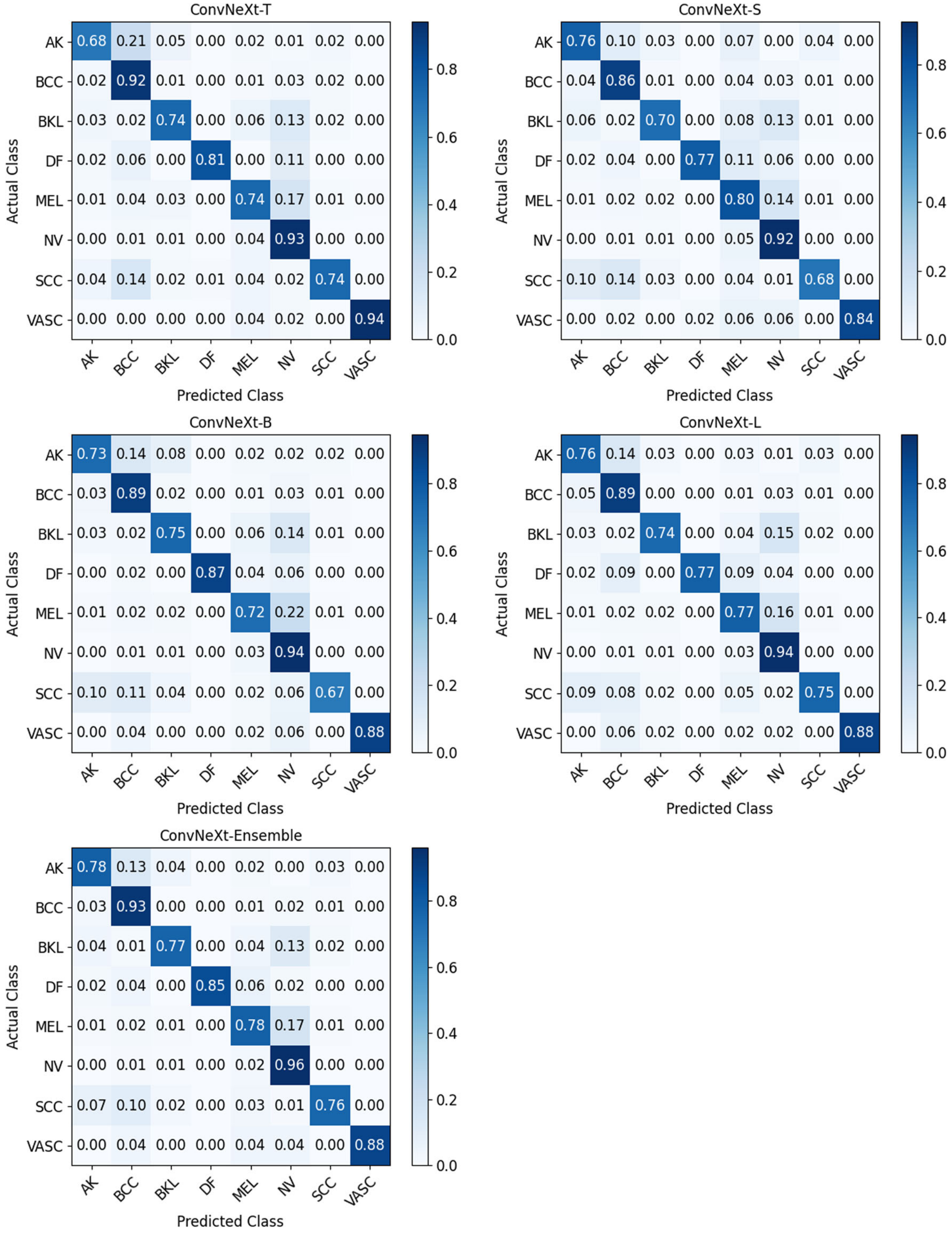

Fig. 3 Confusion matrices depicting the performance on the test set for the eight-class skin lesion classification, highlighting individual sub-versions of ConvNeXt (T: tiny, S: small, B: base, L: large) and the proposed ensemble model, with results specifically from fold 1 test set. The diagonal values represent the sensitivity values for each class

racy with the individual ConvNeXt-T model, which was also improved to 78% with the ensemble model. It is noteworthy that the individual accuracies of the models vary significantly for each class. For instance, the individual performances of tiny, small, base and large models for AK class are 68%, 76%, 73%, and 76%, respectively. In addition, the ensemble

Table 4 Performance comparison of the individual and ensemble of ConvNeXt models

Model	Sensitivity	Specificity	Accuracy
ConvNeXt T	80.5± 0.5867	97.4± 0.0946	96.9± 0.0923
ConvNeXt S	80.2± 2.1921	97.5 ± 0.0672	96.9± 0.0742
ConvNeXt B	80.8 ± 0.7194	97.4 ± 0.0912	97 ± 0.1286
ConvNeXt L	81.3 ± 0.6173	97.5 ± 0.0768	97.2 ± 0.1196
ConvNeXt T-S	82.8 ± 1.6220	97.8 ± 0.0845	97.4 ± 0.0759
ConvNeXt T-B	82.8 ± 0.4938	97.7 ± 0.0893	97.4 ± 0.1403
ConvNeXt T-L	82.6 ± 0.3386	97.7 ± 0.0506	97.5 ± 0.0704
ConvNeXt S-B	82.7 ± 1.2465	97.8 ± 0.0892	97.5 ± 0.1052
ConvNeXt S-L	83.0 ± 1.0438	97.8 ± 0.0804	97.5 ± 0.1018
ConvNeXt B-L	83.1± 0.8498	97.7 ± 0.0661	97.5±0.1575
ConvNeXt T-S-B	83.9 ± 0.7566	97.9± 0.1130	97.6± 0.1293
ConvNeXt T-S-L	83.9 ± 0.8369	97.9 ± 0.0844	97.6 ± 0.0952
ConvNeXt T-B-L	83.8± 0.3986	97.8± 0.0847	97.6±0.1420
ConvNeXt S-B-L	83.6 ±0.5586	97.9 ± 0.0807	97.6 ±0.1070
ConvNeXt T-S-B-L	**84.2±0.6460**	**97.9±0.0852**	**97.7±0.1134**

Table 5 Performance comparison of the ConvNeXt models with state-of-the-art models

Models	Sensitivity	Specificity	Accuracy
Molina et al. [15]	66.5	97.9	97.4
Zhao et al. [16]	68.2	–	–
EfficientNets [17]	72.5	–	–
Kassem et al. [18]	79.8	97.0	94.9
CSLNet [19]	89.6	97.6	89.6
Swin-T [20]	80.3	97.3	96.3
Swin-S [20]	83.0	97.5	96.5
Swin-B [20]	84.5	97.8	96.9
Swin-L [20]	83.1	97.7	97.0
Proposed Method	**84.2**	**97.9**	**97.7**

model achieves 78% accuracy for the AK class and demonstrates higher accuracy than the individual models for almost all other classes as well.

5 Conclusion

Automatic classification of skin lesions is a very challenging step due to various factors such as similar appearance lesions, diverse skin structures, variations in lesion stages, limited or inaccurate data, and artifacts present in dermoscopy images. In this study, we conducted an analysis and comparison of different versions of pre-trained and fine-tuned ConvNeXt models, i.e. Tiny, Small, Base, and Large, for skin lesion classification on publicly available ISIC 2019 dataset. However, the true strength of our approach lies in the ensemble model, which combines all four ConvNeXt models to produce more accurate result. The proposed ensemble model achieved an impressive overall classification accuracy of 97.7%, surpassing the performance of both individual models and state-of-the-art methods. Furthermore, our proposed method yielded a sensitivity value of 84.2% and a specificity value of 97.9%, indicating its ability to accurately classify skin lesions from dermoscopy images. These results highlight the effectiveness of the ConvNeXt architecture and its ensemble approach in addressing the challenges associated with skin lesion classification. The successful application of ConvNeXt models in this study opens up possibilities for developing more robust and reliable automated systems for skin lesion analysis. Future research can explore further enhancements to the ConvNeXt architecture and ensemble learning techniques to improve the performance and generalizability of skin lesion classification systems. Ultimately, such advancements can contribute to early detection, timely treatment, and improved outcomes for patients with skin diseases.

Author Contributions EBK and SA have made equal contributions to the concept and design of the paper.

Funding No funding is received for this work.

Data Availability The dataset proposed in this research is available to download from the following link: https://challenge.isic-archive.com/data/#2019.

Declarations

Conflict of interest The authors declare that they have no Conflict of interest.

Ethical Approval Not applicable.

References

1. Dildar, M., Akram, S., Irfan, M., Khan, H.U., Ramzan, M., Mahmood, A.R., et al.: Skin cancer detection: a review using deep learning techniques. Int. J. Environ. Res. Public Health **18**(10), 5479 (2021)
2. WHO.: Radiation: Ultraviolet (UV) radiation and skin cancer. (2023). Available from: https://www.who.int/news-room/questions-and-answers/item/radiation-ultraviolet-(uv)-radiation-and-skin-cancer
3. Rundo, F., Banna, G.L., Conoci, S.: Bio-inspired deep-CNN pipeline for skin cancer early diagnosis. Computation **7**(3), 44 (2019)
4. Warszawik-Hendzel, O., Olszewska, M., Maj, M., Rakowska, A., Czuwara, J., Rudnicka, L.: Non-invasive diagnostic techniques in the diagnosis of squamous cell carcinoma. J. Dermatol. Case Rep. **9**(4), 89 (2015)
5. Vesal, S., Patil, S.M., Ravikumar, N., Maier, A.K.: A Multi-task Framework for Skin Lesion Detection and Segmentation. arXiv:1808.01676
6. Kavitha, P., Jayalakshmi, V.: Comparative Study of DNN Models for Skin Cancer Detection. In: 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE; pp. 1350–1356 (2022)
7. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., et al.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)
8. Chen, X., Wang, X., Zhang, K., Fung, K.M., Thai, T.C., Moore, K., et al.: Recent advances and clinical applications of deep learning in medical image analysis. Medical Image Analysis, pp. 102444 (2022)
9. Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., et al.: Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. J. Big Data **8**, 1–74 (2021)
10. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929. (2020)
11. Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S.: A convnet for the 2020s. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; pp. 11976–11986 (2022)
12. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision; pp. 10012–10022 (2021)
13. Liew, W.S., Tang, T.B., Lin, C.H., Lu, C.K.: Automatic colonic polyp detection using integration of modified deep residual convolutional neural network and ensemble learning approaches. Comput. Methods Programs Biomed. **206**, 106114 (2021)
14. Younas, F., Usman, M., Yan, W.Q.: A deep ensemble learning method for colorectal polyp classification with optimized network parameters. Appl. Intell. **53**(2), 2410–2433 (2023)
15. Molina-Molina, E.O., Solorza-Calderón, S., Álvarez-Borrego, J.: Classification of dermoscopy skin lesion color-images using fractal-deep learning features. Appl. Sci. **10**(17), 5954 (2020)
16. Zhao, C., Shuai, R., Ma, L., Liu, W., Hu, D., Wu, M.: Dermoscopy image classification based on StyleGAN and DenseNet201. IEEE Access **9**, 8659–8679 (2021)
17. Gessert, N., Nielsen, M., Shaikh, M., Werner, R., Schlaefer, A.: Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data. MethodsX. **7**, 100864 (2020)
18. Kassem, M.A., Hosny, K.M., Fouad, M.M.: Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning. IEEE Access. **8**, 114822–114832 (2020)
19. Iqbal, I., Younus, M., Walayat, K., Kakar, M.U., Ma, J.: Automated multi-class classification of skin lesions through deep convolutional neural network with dermoscopic images. Comput. Med. Imaging Graph. **88**, 101843 (2021)
20. Ayas, S.: Multiclass skin lesion classification in dermoscopic images using swin transformer model. Neural Comput. Appl. **35**(9), 6713–6722 (2023)
21. Tschandl, P., Rosendahl, C., Kittler, H.: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci. Data. **5**(1), 1–9 (2018)
22. Codella, N.C., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., et al.: Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE; pp. 168–172 (2018)
23. Combalia, M., Codella, N.C., Rotemberg, V., Helba, B., Vilaplana, V., Reiter, O., et al.: Bcn20000: Dermoscopic lesions in the wild. arXiv preprint arXiv:1908.02288. (2019)
24. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., et al.: Dermatologist-level classification of skin cancer with deep neural networks. Nature. **542**(7639), 115–118 (2017)
25. Abbas, Q., Celebi, M.E.: DermoDeep-a classification of melanoma-nevus skin lesions using multi-feature fusion of visual features and deep neural network. Multimedia Tools Appl. **78**(16), 23559–23580 (2019)
26. Pacheco, A.G., Krohling, R.A.: An attention-based mechanism to combine images and metadata in deep learning models applied to skin cancer classification. IEEE J. Biomed. Health Inform. **25**(9), 3554–3563 (2021)
27. Zahira, S., Abbasb, A.W., Khanc, R.U., Ullahd, M.: Vision sensor assisted fire detection in IoT environment using ConvNext. J. Artifi. Intell. Syst. **5**, 23–35 (2023)
28. Tian, G., Wang, Z., Wang, C., Chen, J., Liu, G., Xu, H., et al.: A deep ensemble learning-based automated detection of COVID-19 using lung CT images and Vision Transformer and ConvNeXt. Front. Microbiol. **13**, 96 (2022)
29. Combalia, M., Codella, N., Rotemberg, V., Carrera, C., Dusza, S., Gutman, D., et al.: Validation of artificial intelligence prediction models for skin cancer diagnosis using dermoscopy images: the 2019 International Skin Imaging Collaboration Grand Challenge. The Lancet Digital Health. **4**(5), e330–e339 (2022)
30. : The International Skin Imaging Collaboration. Accessed Feb. 6, Available from: https://www.isic-archive.com/ (2024)
31. Kohavi, R., Provost, F.: Glossary of terms. Appl. Mach. Learn. Knowl. Discov. Process. **30**, 96 (1998)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.