electronics

MDPI

Article

A Skin Disease Classification Model Based on DenseNet and ConvNeXt Fusion

Mingjun Wei [1], Qiwei Wu [1], Hongyu Ji [2], Jingkun Wang [3], Tao Lyu [4], Jinyun Liu [1,*] and Li Zhao [3,*]

[1] College of Artificial Intelligence, North China University of Science and Technology, Tangshan 063210, China
[2] School of Biosciences, University of Sheffield, Sheffield S10 2TN, UK
[3] Beijing National Research Center for Information Science and Technology, Institute for Precision Medicine, Tsinghua University, Beijing 100084, China
[4] Department of Obstetrics and Gynecology, Beijing Tsinghua Changgung Hospital, Beijing 102218, China
* Correspondence: liujy23@ncst.edu.cn (J.L.); zhaoli@tsinghua.edu.cn (L.Z.)

Abstract: Skin disease is one of the most common diseases. Due to the intricate categories of skin diseases, their symptoms being very similar in the early stage, and the lesion samples being extremely unbalanced, their classification is challenging. At the same time, under the conditions of limited data, the generalization ability of a single reliable convolutional neural network model is weak, the feature extraction ability is insufficient, and the classification accuracy is low. Therefore, in this paper, we proposed a convolutional neural network model for skin disease classification based on model fusion. Through model fusion, deep and shallow feature fusion, and the introduction of an attention module, the feature extraction capacity of the model was strengthened. In addition, a series of works such as model pre-training, data augmentation, and parameter fine-tuning were conducted to upgrade the classification performance of the model. The experimental results showed that when working on our private dataset dominated by acne-like skin diseases, our proposed model outperformed the two baseline models of DenseNet201 and ConvNeXt_L by 4.42% and 3.66%, respectively. On the public HAM10000 dataset, the accuracy and f1-score of the proposed model were 95.29% and 89.99%, respectively, which also achieved good results compared with other state-of-the-art models.

Keywords: attention module; classification; feature fusion; model fusion; skin disease

Citation: Wei, M.; Wu, Q.; Ji, H.; Wang, J.; Lyu, T.; Liu, J.; Zhao, L. A Skin Disease Classification Model Based on DenseNet and ConvNeXt Fusion. *Electronics* **2023**, *12*, 438. https://doi.org/10.3390/electronics12020438

Academic Editor: Maria Evelina Fantacci

Received: 1 December 2022
Revised: 4 January 2023
Accepted: 12 January 2023
Published: 14 January 2023

1. Introduction

Skin disease is a severe global public health problem that affects a large number of people [1]. The symptoms of skin diseases are diverse, and the changing of the symptoms is a long-term process. It is difficult for ordinary people to determine the type of skin disease with the naked eye, and most people often neglect the changes in their skin symptoms, which can lead to severe consequences such as permanent skin damage and even the risk of skin cancer [2]. In addition, the early treatment of skin cancer can decrease morbidity and mortality [3].

In addition, due to the rapid development of deep learning technology, it has rapidly become the preferred method for medical image analysis [4,5]. In addition, compared with traditional classification methods, deep learning has a stronger robustness and a better generalization ability [6]. In the meantime, convolutional neural networks are one of the most well-known and representative deep learning models [7,8]. It has been widely used in many aspects of medical image analysis [9,10], and great progress has been made in medical image classification. For example, Datta. et al. [11] combined soft-attention and Inception ResNet-V2 [12] (IRv2) to construct an IRV2-SA model for dermoscopic image classification. This combination improved the sensitivity score compared to the baseline model, reaching 91.6% on the ISIC2017 [13] dataset. Apart from that, its accuracy on the HAM10000 [14] dataset was 93.7%, which was 4.7% higher than the baseline model. Lan. et al. [15] proposed a capsule network method called FixCaps. It is an improved convolutional neural network

model based on CapsNets [16] with a larger receptive domain. It works by applying a high-performance large kernel with a kernel size of up to 31 × 31 at the bottom convolutional layer. At the same time, an attention mechanism was introduced to reduce the loss of spatial information caused by convolution and pooling, and it achieved an accuracy of 96.49% and an f1-score of 86.36% on the HAM10000 dataset.

The IRV2-SA model and FixCaps model perform well in terms of classification accuracy. However, they are not impeccable in terms of other classification performance evaluation criteria, and the classification performance is not satisfactory in classifications with a restricted individual sample data. Enhancing their classification accuracy is problematic because of the restricted available image data of skin diseases and the extreme imbalance of lesion samples. In addition, the categories of skin diseases are elaborate, and the symptoms are very analogous in the early stages, which causes the model classification to be more problematic. At the same time, the generalization ability of a single reliable network model qualified with restricted data is weak, and the feature extraction ability is insufficient. Attaining a high classification accuracy is still challenging. The common research strategy to solve the problem of small data samples and class imbalance is data augmentation or enhancing the feature extraction ability of the model.

All in all, the main contributions of this paper can be summarized in the following points:

1. In this work, a convolutional neural network (CNN) model based on model fusion was proposed for skin disease classification. DenseNet201 [17] and ConvNeXt_L [18] were selected as the backbone sub-classification models for the model fusion.
2. To enhance the feature extraction ability of the proposed network model, the Efficient Channel Attention [19] module and the Gated Channel Transformation [20] attention module were introduced into the core blocks of DenseNet201 and ConvNeXt_L, respectively.
3. A parallel strategy was applied to fuse the features of the deep and shallow layers to further enhance the feature-extraction ability of the model.
4. The classification performance of the model was improved through a series of works such as model pre-training, data augmentation, and parameter fine-tuning.
5. Extensive experiments were conducted to compare the proposed model with the basic CNN models commonly used in recent years to ensure the validity of this work. The experiments were carried out by the proposed network model on a private dataset dominated by acne-like skin diseases, and training and testing were conducted on the public HAM10000 [14] (Human-Against-Machine with 10000 training images) dataset with an extreme imbalance in skin diseases, and the proposed model was compared with other state-of-the-art models on the HAM10000 dataset. This verified the generalization capacity and the accuracy of the proposed network model.

2. Related Work

CNN models have been widely explored for skin disease classification, and some of these models have achieved very good classification performances. Below, we summarized the relevant published work of some researchers in the field of skin disease image classification.

Many researchers have proposed reliable multi-class CNN models. Mobiny et al. [21] proposed an approximate risk-aware deep Bayesian model named Bayesian DenseNet-169, which outputs an estimate of model uncertainty without additional parameters or significant changes to the network architecture. It increased the classification accuracy of the base DenseNet169 [17] model from 81.35% to 83.59% on the HAM10000 dataset. Wang et al. [22] propose an interpretability-based CNN model. It is a multi-class classification model that takes skin lesion images and patient metadata as the input for skin lesion diagnosis. It achieved a 95.1% and 83.5% accuracy and sensitivity, respectively, on the HAM10000 dataset. Allugunti et al. [23] created a multi-class CNN model for diagnosing skin cancer. The proposed model makes a distinction between lesion maligna, superficial spreading, and

nodular melanoma. This permits the early diagnosis of the virus and the quick isolation and therapy necessary to stop the further transmission of infection. Anand et al. [24] modified the Xception [25] model by adding layers such as a pooling layer, two dense layers, and a dropout layer. A new fully connected (FC) layer changed the original FC layer with seven skin disease classes. It had a classification accuracy of 96.40% on the HAM10000 dataset.

Improving the classification accuracy of the model by using ensemble learning is also an effective method. Thurnhofer-Hemsi et al. [26] proposed an ensemble composed of improved CNNs combined with a regularly spaced test-time-shifting technique for skin lesion classification. It builds up multiple test input images via a shift technique and passes it to each classifier passed to the ensemble and then combines all the outputs for classification. It had a classification accuracy of 83.6% on the HAM10000 dataset.

Through the introduction of an attention module, the feature extraction ability of a model can be enhanced, thereby improving the classification performance of the model. Karthik et al. [27] replaced the standard Squeeze-and-Excite [28] block in the Efficient-NetV2 [29] model with an Efficient Channel Attention [19] block, and the total number of training parameters dropped significantly. The test accuracy of the model reached 84.70% in four types of skin disease datasets including acne, actinic keratosis, melanoma and psoriasis.

Through image processing techniques such as image conversion, equalization, enhancement and segmentation, the accuracy of image classification can be enhanced. Abayomi-Alli et al. [30] propose an improved data augmentation model for the effective detection of melanoma skin cancer. The method was based on oversampling data embedded in a nonlinear low-dimensional manifold to create synthetic melanoma images. It achieved a 92.18%, 80.77%, 95.1% and 80.84% accuracy, sensitivity, specificity and f1-score, respectively, on the PH2 [31] dataset. Hoang et al. [32] proposed a novel method using a new segmentation approach and wide-ShuffleNet for skin lesion classification. It first separates the lesion from the background by computing an entropy-based weighted sum first-order cumulative moment (EW-FCM) of the skin image. The segmentation results are then input into a new deep learning structure, wide-ShuffleNet, and classified. It achieved a 96.03%, 70.71%, 75.15%, 72.61% and 84.80% specificity, sensitivity, precision, f1-score and accuracy, respectively, on the HAM10000 dataset. Malibari et al. [33] proposed an Optimal Deep-Neural-Network-Driven Computer-Aided Diagnosis Model for their skin cancer detection and classification model. The model primarily applies a Wiener-filtering-based pre-processing step followed by a U-Net segmentation approach. The model achieved a maximum accuracy of 99.90%. Nawaz et al. [34] proposed an improved Deep-Learning-based method, namely, the DenseNet77-based UNET model. Their experiments demonstrated the robustness of the model and its ability to accurately identify skin lesions of different colors and sizes. It obtained a 99.21% and 99.51% accuracy on the ISIC2017 [13] and ISIC2018 [35] datasets, respectively.

Therefore, by summarizing the related work published by these researchers in the field of skin disease image classification, we proposed a CNN model for skin disease classification based on model fusion. In addition, through a series of work such as model fusion, deep and shallow feature fusion, the introduction of an attention module, model pre-training, data augmentation and parameter fine-tuning, the classification performance of the proposed model was enhanced.

3. Method

First, we trained and tested the classification performance of basic CNN models (including ResNet50 [36], EfficientNet_B4 [37], DenseNet201 [17] and ConvNeXt_L [18]) that have been commonly used in recent years on our private dataset dominated by acne-like skin diseases. This was a typical dataset with a small amount of sample data and extremely unbalanced categories. Then, after the research, it was found that the two CNN models DenseNet201 and ConvNeXt_L achieved a good classification performance, and their accuracy rates were 92.12% and 92.88%, respectively, which were the top two

best-performing models. Multi-model fusion can be configured with any number of sub-classification CNN models at the same time. However, the more sub-classifiers there are, the less computationally efficient the model is, and it is important to strike a balance [38]. Therefore, we chose DenseNet201 and ConvNeXt_L as the backbone sub-classification models of our model fusion.

3.1. Improving the DenseNet Model

DenseNet [17] is a classic image classification model that proposes a more aggressive dense connection mechanism, that is, for each layer, the feature-maps of all the preceding layers are used as inputs, and its own feature maps are used as inputs into all subsequent layers. It has four sub-versions, and we improved theDenseNet201 model with deeper network layers. The dense connection mechanism is helpful for feature reuse in the network, effectively slowing down the gradient disappearance problem, and it can better complete most image classification tasks. Huang et al. [17] demonstrated the robustness of the architecture using the vanishing gradient problem.

At the same time, an attention mechanism can improve a network model's ability to extract image features, help the network to obtain a region of interest, reduce the attention paid to non-important information and improve the network's classification performance [39]. The most representative attention mechanism is SENet [28]. At the heart of SENet is a squeeze–excitation block that is used to collect global information, capture channel relationships and improve the representation capability of the model. However, we introduced a squeeze–excitation block into the internal module of DenseNet, which did not improve the classification performance satisfactorily on our private dataset dominated by acne-like skin diseases. In addition, to avoid higher model complexity, SENet reduces the number of channels. However, this does not directly model the correspondence between weight vectors and inputs, resulting in a poor classification performance improvement. To improve this shortcoming, Wang et al. [19] proposed the Efficient Channel Attention (ECA) block, which utilizes a 1D convolution to determine the interactions between channels. It includes a squeeze module for aggregating global spatial information and an effective excitation module for modeling cross-channel interactions. It controls the model complexity by considering only direct interactions between each channel and its k-nearest neighbors rather than indirect correspondence.

Therefore, inspired by the characteristics of both, we introduced the ECA block into the internal module of DenseNet to form a new block structure. It can be represented by Equation (1), where layer l receives all the feature mappings from the previous layers, $x_0,\ x_1, \ldots, x_{l-1}$ are the inputs, $[x_0, x_1, \ldots, x_{l-1}]$ refers to the concatenation of feature mappings generated in layer $0, 1, \ldots, l-1$ and H_l represents the non-linear transformation function, which is a combination operation that consists of a series of batch normalizations (BN), a rectified linear unit (ReLU), an ECA block, pooling and convolution (Conv). Its improved block is composed of multiple improved layers through dense links, and the structure of the improved layer and the structure of the improved block are shown in Figure 1. Our improved layer consisted of BN, ReLU, 1×1 Conv, BN, ReLU, 3×3 Conv and an ECA block from top to bottom.

$$x_l = H_l([x_0, x_1, \ldots, x_{l-1}]) \tag{1}$$

3.2. Improved the ConvNeXt Model

ConvNeXt [18] was used to adjust the existing classic ResNet [36] model and introduce some of the latest ideas and technologies of the Swin Transformer [40] model into the existing modules to improve the classification performance of the model. It has five sub-versions, and we improved the ConvNeXt_L model with deeper network layers. The main backbone network is composed of 4 different stages, each stage being composed of several blocks, and ConvNeXt was used to adjust the ratio of blocks in each stage to 1:1:3:1. It replaced the 3×3 convolution with a 3×3 depthwise convolution and increased

the number of base channels from 64 to 96. Then, an inverted bottleneck structure was adopted, while the rectified linear unit (ReLU) and batch normalization (BN) were replaced by a Gaussian error linear unit [41] (GELU) and layer normalization [42] (LN). Finally, the convolution kernel was enlarged to 7×7.

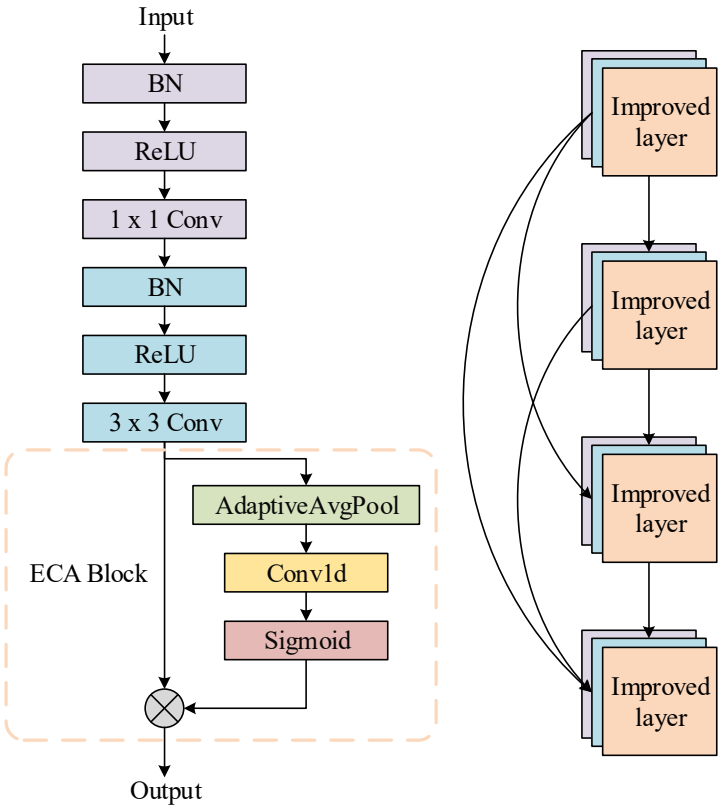


Figure 1. The structure of the improved DenseNet block.

Similarly, we introduced an attention module, namely the ECA block, to the core structure of ConvNeXt. However, on our private dataset dominated by acne-like skin diseases, the combination of ECA blocks with ConvNeXt did not significantly improve the classification performance. Therefore, we combined it with a new gated-channel transformation [20] (GCT) attention module. The GCT module can efficiently gather information while explicitly modeling channel-wise relationships. It takes a normalized approach to creating competition or partnership between channels. Meanwhile, it designs a global contextual embedding operator and controls the weights of each channel before normalization, thus making GCT learnable. GCT first computes the l_2-norm of each channel to collect global information. The features are then scaled with a learnable vector, α, and channel-normalized (CN), and then the normalization is rescaled with a learnable bias, β, and a scale parameter, γ. It then adopts tanh activation to regulate the attention carrier. Eventually, it multiplies the input by the attention vector while adding an identity connection. In addition, it is lightweight; even if it is added after each convolutional layer of the model, the computational demands will not shatter. Finally, the core block structure of our improved ConvNeXt model is shown in Figure 2.

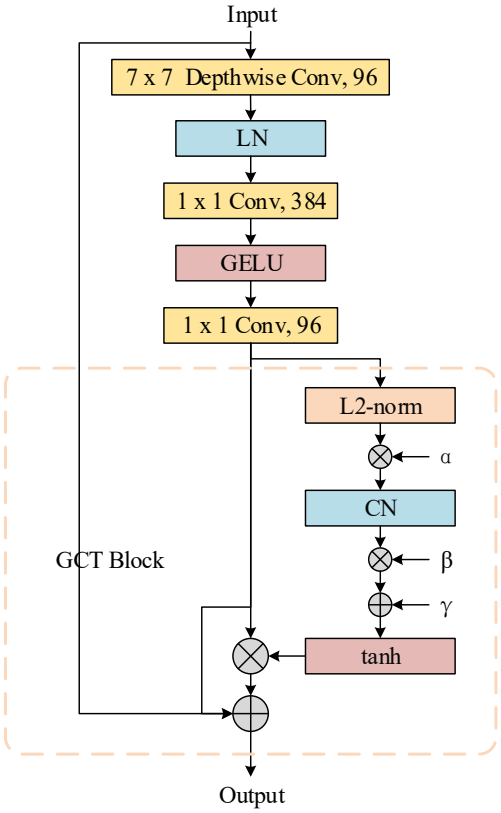


Figure 2. The structure of the improved ConvNeXt block.

3.3. Macro Design

Different sub-models have different expressive abilities, and by combining the parts they are good at, a model that is "accurate" in all aspects is obtained. Therefore, we fused the two improved sub-models to form the backbone of our classification model.

The features extracted by the shallow network were relatively close to the input and contained more pixel information, that is, fine-grained information such as the color, texture, edges and corners of the image. The receptive field of the shallow network was smaller, and the overlapping area of the receptive field was also smaller, so the shallow network could capture more details. However, the semantics were lower due to less convolution going through. The features extracted by the deep network were closer to the output and contained more abstract information, that is, coarse-grained information such as semantic information. However, the resolution was low, and the perception of details was poor. Therefore, combining the characteristics of the two, a parallel strategy was adopted to fuse the deep and shallow features. It can be represented by Function (2), where x represents the input, Conv represents the 2×2 convolution operation with stride 2, and Dropout represents the operation of randomly ignoring some features, which could significantly reduce the overfitting phenomenon [43].

$$G(x) = Conv(Dropout(x)) \tag{2}$$

The complete structure of our proposed model is shown in Figure 3. For the improved DenseNet model, the features output from the second block are first passed through the (2) operation and then added and fused with the features output from the third block. The fused features are again subjected to the (2) operation, and they are then added and fused with the features output by the fourth block to serve as the final output features. The extracted features are first adaptively average-pooled, and then the multi-dimensional features are one-dimensionalized by the flattening layer. For the improved ConvNeXt model, the features output by the third stage are first subjected to the (2) operation, and

they are then added and fused with the features output by the fourth stage as the final output features. The extracted features are adaptively average-pooled. Finally, the features output by the two improved sub-models are concatenated for classification. In addition. all the models were pre-trained on ImageNet [44], where the weight files were either obtained from Torchvision or Github. In order to match our proposed model, we replaced and deleted some keys in the weight file.

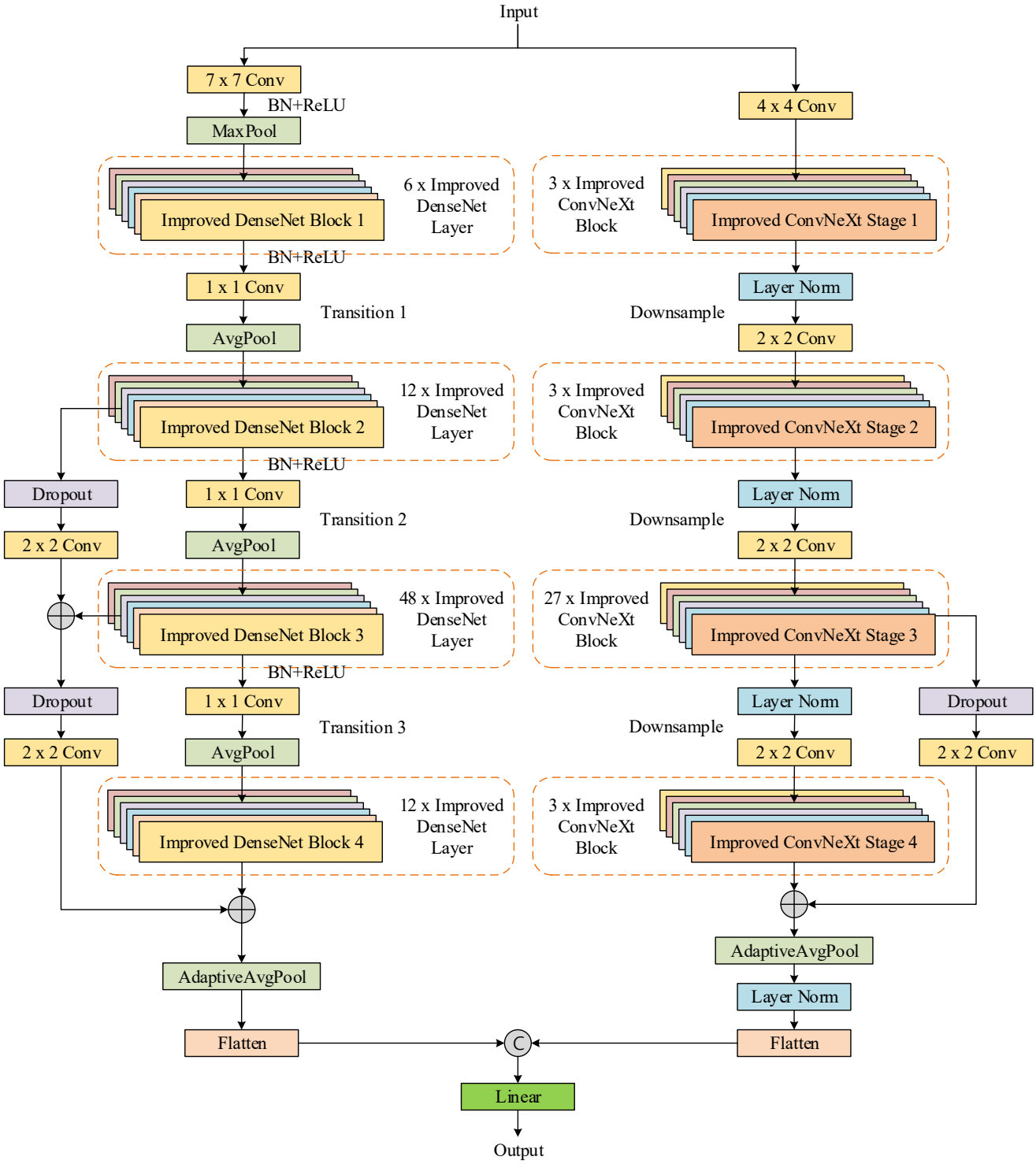
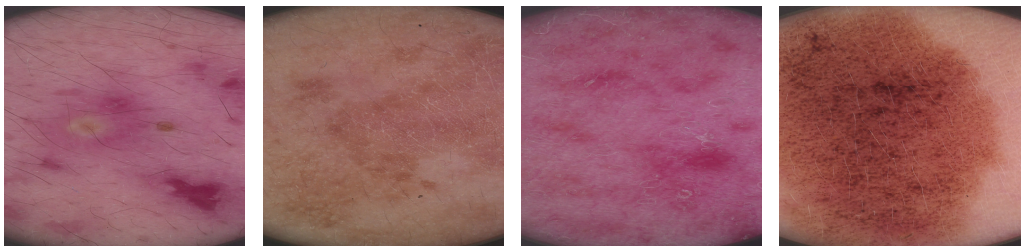


Figure 3. The full structure of the proposed model.

4. Experient and Results

4.1. Datasets

The first experimental dataset in this paper was provided by Peking Union Medical College Hospital, and all participants provided informed consent. This dataset had a total of 2600 images, including 1600 images of acne skin diseases, 400 images of melasma skin diseases, 300 images of rosacea skin diseases and 300 images of nevus of Ota skin diseases. These images and labels were rigorously reviewed by multiple experienced dermatologists. Some of the sample images from the datasets are shown in Figure 4. We randomly divided the dataset into a training set and a test set according to a ratio of 8:2. The fact that there were far more acne skin disease images than the other three classes led to an irregular distribution of skin disease images and an unbalanced dataset. Therefore, we used data augmentation to balance the data so as to improve the classification performance of the model, reduce the overfitting of the data and make the model more stable in the learning process [45]. We expanded the training set eight times by horizontal flipping, vertical flipping, increasing the brightness, center cropping, Cutout [46], Cutmix [47], Augmix [48] and Random Erasing [49], but we did not modify the test set. Figure 5 shows the number of images for each class of skin disease in the test set. Before training, we normalized the pixel values of the input images to a [0, 1] range and resized the images to 512 × 512 pixels.



(a) Acne (b) Melasma (c) Rosacea (d) Nevus of Ota

Figure 4. Sample images of the four skin diseases.

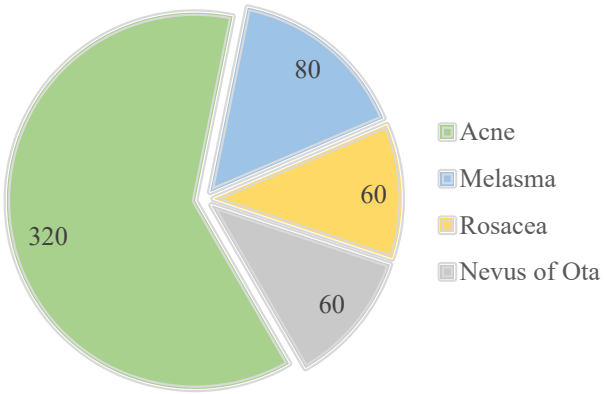


Figure 5. The number of images of each class of skin disease in the test set.

In addition, in order to verify the generalization ability of our proposed network model and make the accuracy of the model more convincing, we conducted additional experiments on the public dataset HAM10000 [14] (Human-Against-Machine with 10000 training images). It contains 10015 images of skin diseases that are divided amongst seven classes, including three hundred and twenty-seven images of actinic keratosis and intraepithelial carcinoma (AKIEC), five hundred and fourteen images of basal cell carcinoma (BCC), one thousand and ninety-nine images of benign keratosis-like lesions (BKL), one hundred and fifteen images of dermatofibroma (DF), one thousand one hundred and thirteen images of melanoma (MEL), six thousand seven hundred and five images of melanocytic nevi (NV) and one hundred and forty-two images of vascular skin lesions (VASC). So, it is a dataset with extremely imbalanced skin disease classes. Some sample images from the

HAM10000 dataset are shown in Figure 6. Then, we normalized the dataset to a uniform size (300 × 300). For a fair comparison with the other models, we divided the dataset in two ways. In the first way, 828 skin disease images were randomly extracted as the test set, which was the same as the dataset division of the models IRv2-RA [11], FixCaps [15], etc. In the second way, we randomly divided the training set and the test set according to a ratio of 8:2. The test set had 2000 skin disease images, which was the same as the dataset division of the models Shifted2-Nets [26], etc. Table 1 shows the number of images for each class of skin disease in the test set for the two partitions. In addition, in order to make the model have better experimental results, the training dataset was processed with the same data augmentation method as the first private dataset.

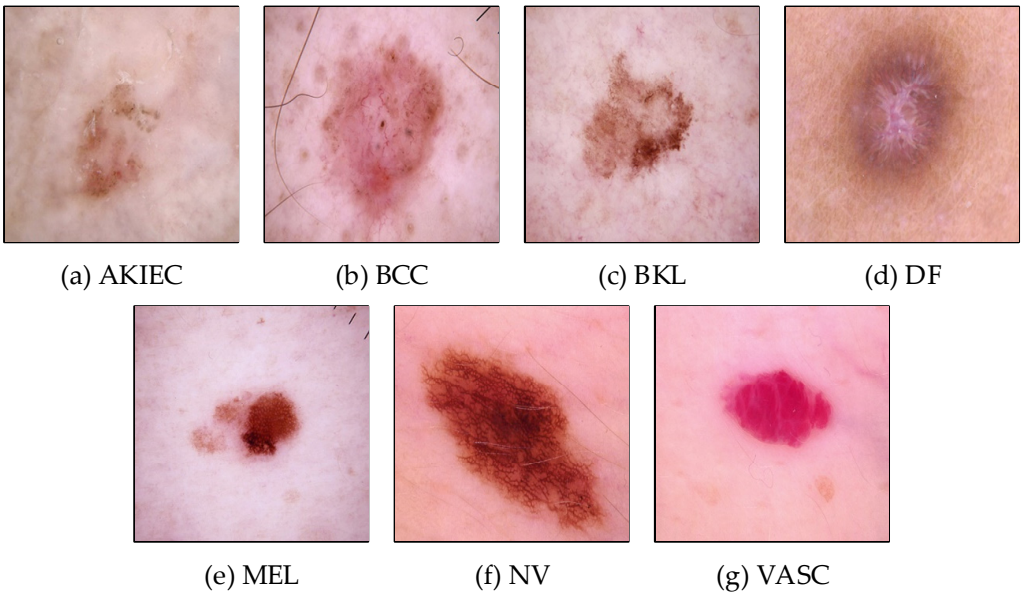


Figure 6. Sample images of the seven skin diseases.

Table 1. The number of images for each class of skin disease in the test set for the two partitions.

Class	First Way	Second Way
AKIEC	23	65
BCC	26	103
BKL	66	219
DF	6	23
MEL	34	221
NV	663	1341
VASC	10	28
Total	828	2000

4.2. Metrics

In the conducted experiments, various metrics were used to evaluate the performance of the proposed models and compared it to that of four basic models, namely ResNet50, EfficientNet_B4, DenseNet201 and ConvNeXt_L. We also compared the models proposed by others. The preliminary metrics were accuracy, precision, recall and f1-score. To extend our metrics to multiclass classification, the macro-average was also calculated.

Accuracy is the most intuitive performance measure, and it is simply a ratio of the correctly predicted observations to total observations. The accuracy was calculated by using (3) [50], where TP (true positives) represents the correctly predicted positive values, which means that the value of the actual class is yes and the value of the predicted class is also yes. TN (true negatives) represents the correctly predicted negative values, which means that the value of the actual class is no and the value of the predicted class is also

no. FP (false positives) represents when the actual class is no and the predicted class is yes. FN (false negatives) represents when the actual class is yes but the predicted class in no.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. The precision was calculated using Equation (4) [51].

$$precision = \frac{TP}{TP + FP} \tag{4}$$

Recall is the proportion of actual positives that are identified correctly. The recall was calculated using Equation (5) [51].

$$recall = \frac{TP}{TP + FN} \tag{5}$$

The f1-score takes into account both precision and recall. The f1-score was calculated using Equation (6) [51].

$$f_1 - score = \frac{2 \times precision \times recall}{precision + recall} \tag{6}$$

The macro-average treats each class equally, with all classes having the same weight. It is obtained by adding up the evaluation metrics (precision/recall/f1-score) of different classes and calculating the average. For example, to calculate the macro-average of the metric precision of k-class, its macro-average is calculated by using (7) [52].

$$precision_{macro-\text{average}} = \frac{precision_1 + precision_2 + \ldots + precision_k}{k} \tag{7}$$

4.3. Results

We conducted experiments on both our private dataset and the public dataset HAM10000. The operating system of the experimental server was Ubuntu20.04, which was configured with 1 AMD EPYC 7642 48-Core CPU and 8 NVIDIA RTX 3090 24GB GPUs.

4.3.1. The First Dataset

The experimental environment used was built based on the deep learning framework pytorch1.10. The stochastic gradient descent (SGD) algorithm [53] was used to optimize the model; the initial learning rate was 0.01, the momentum was 0.9, the weight decay was 0.0001 and the batch size was 64. The MultiStepLR algorithm was used to dynamically adjust the learning rate and reduce the learning rate in the 10th, 15th and 25th epoch, respectively, and the gamma was 0.1. Categorical cross-entropy was selected as the loss function. All models werre trained for 60 epochs.

First of all, after 60 rounds of training for all the models, the best test set accuracies of each model were obtained, as shown in Figure 7, where the best test set accuracies of ResNet50, EfficientNet_B4, DenseNet201, ConvNeXt_L and our proposed model were 88.65%, 91.54%, 92.12%, 92.88% and 96.54%, respectively. These results confirmed that our proposed model outperformed, in terms of accuracy, the other four basic models used in the comparison. More precisely, our proposed model improved the accuracy by 4.42 percentage points and 3.66 percentage points, respectively, compared with the two DenseNet201 and ConvNeXt_L baseline models. Compared with the other models, our proposed model was also significantly improved.

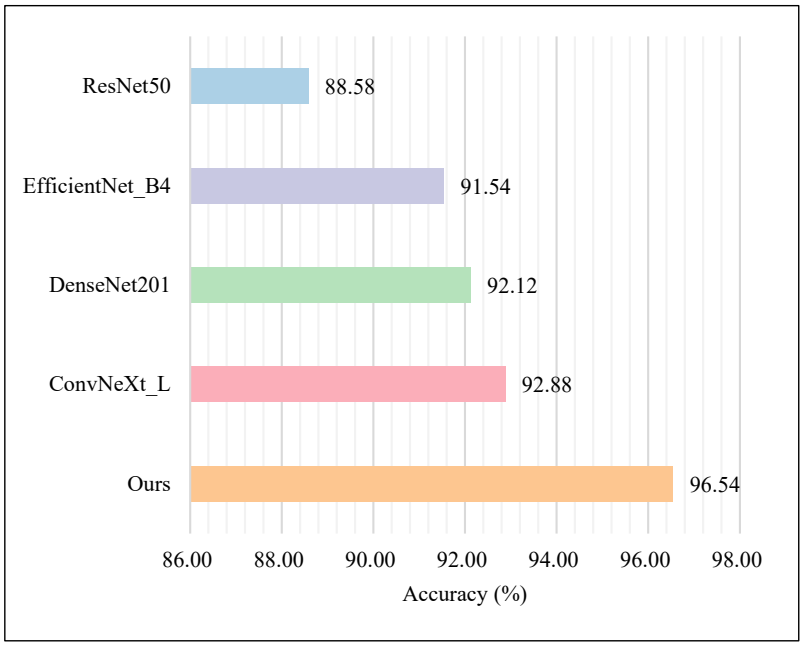


Figure 7. The best test accuracy of each model on our private dataset.

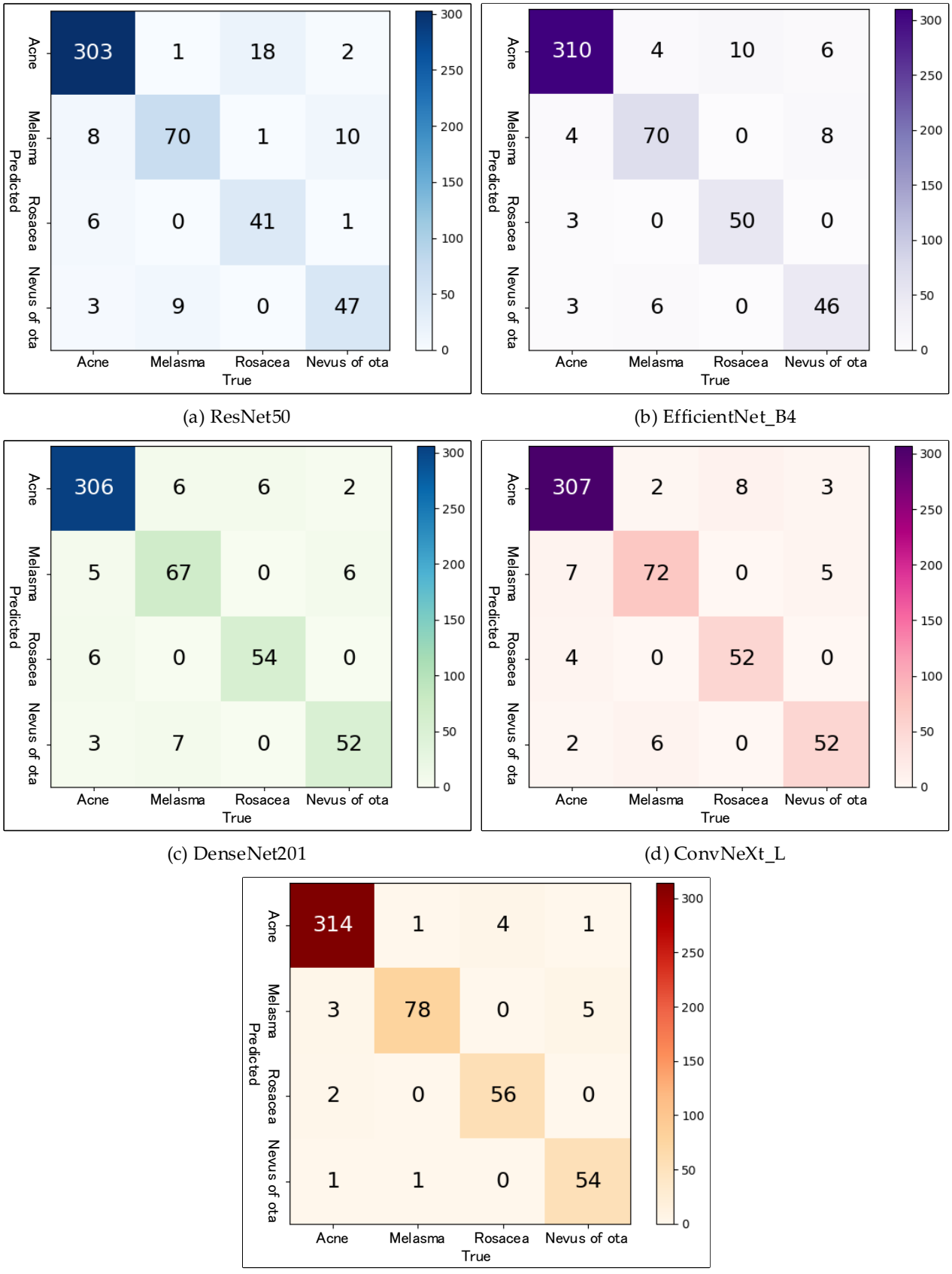
In addition, the confusion matrix corresponding to the best accuracy of each model is shown in Figure 8. Then, these confusion matrixes were used to calculate the precision of each model separately for each class based on (4), as shown in Table 2. Meanwhile, the macro-average of precision was calculated based on (7). Similarly, the recall, f1-score and corresponding macro-average of each model for each class were calculated based on (5), (6) and (7), respectively, as shown in Tables 3 and 4. It can be seen from Tables 2–4 that our proposed model was not only better than the other four basic models in terms of accuracy but also better than the other four basic models in terms of precision, recall and f1-score. At the same time, our proposed model not only outperformed the other models in the categories with more data but also performed better in the categories with less data. For example, our proposed model outperformed ResNet50 by 16.77 percentage points in terms of precision on the images of nevus of Ota skin disease. Finally, these experimental results demonstrated that our proposed model had a better classification performance.

Table 2. The precision of each model on our private dataset; the unit is %.

Model	Acne	Melasma	Rosacea	Nevus of Ota	Macro-Average
ResNet50	93.52	78.65	85.42	79.66	84.31
EfficientNet_B4	93.94	85.37	94.34	83.64	89.32
DenseNet201	95.62	85.90	90.00	83.87	88.85
ConvNeXt_L	95.94	85.71	92.86	86.67	90.30
Ours	98.12	90.70	96.55	96.43	95.45

Table 3. The recall of each model on our private dataset; the unit is %.

Model	Acne	Melasma	Rosacea	Nevus of Ota	Macro-Average
ResNet50	94.69	87.50	68.33	78.33	82.21
EfficientNet_B4	96.88	87.50	83.33	76.67	86.10
DenseNet201	95.62	83.75	90.00	86.67	89.01
ConvNeXt_L	95.94	90.00	86.67	86.67	89.82
Ours	98.12	97.50	93.33	90.00	94.74



(a) ResNet50

(b) EfficientNet_B4

(c) DenseNet201

(d) ConvNeXt_L

(e) Ours

Figure 8. The confusion matrix of each model on our private dataset.

Table 4. The f1-score of each model on our private dataset, the unit is %.

Model	Acne	Melasma	Rosacea	Nevus of Ota	Macro-Average
ResNet50	94.10	82.84	75.93	78.99	82.97
EfficientNet_B4	95.39	86.42	88.49	80.00	87.58
DenseNet201	95.62	84.81	90.00	85.25	88.92
ConvNeXt_L	95.94	87.80	89.66	86.67	90.02
Ours	98.12	93.98	94.91	93.10	95.03

4.3.2. The Second Dataset

Comparing the models on the public dataset HAM10000, the experimental environment of this dataset was basically the same as the experimental environment of the first private dataset. Similarly, categorical cross-entropy was selected as the loss function, the initial learning rate was 0.01, the momentum was 0.9 and the weight decay was 0.0001, but the batch size was 128. The MultiStepLR algorithm was used to dynamically adjust the learning rate and reduce the learning rate in the eighth, fifteenth and twentieth epoch, respectively, and the gamma was 0.1. All of the models were trained for 40 epochs.

To begin with, based on the dataset divided in the first way (test set of 828 images), the accuracy of the four basic models, our proposed model and those proposed by others are shown in Table 5. It can be seen from Table 5 that our proposed model was improved by 2.42 percentage points and 1.57 percentage points, respectively, compared with the two DenseNet201 and ConvNeXt_L baseline models. Compared with CNN [54], IM-CNN [22] and IRv2-RA [11], our proposed model outperformed them in terms of accuracy by 9.31%, 1.89% and 0.19%, respectively, but compared with FixCaps [15], our proposed model was 1.2% lower in terms of accuracy.

Table 5. The accuracy of each model on the public dataset HAM10000 (test set of 828 images); the unit is %.

Model	Accuracy
ResNet50	90.10
EfficientNet_B4	93.36
DenseNet201	92.87
ConvNeXt_L	93.72
IRv2-RA [11]	93.47
FixCaps [15]	96.49
IM-CNN [22]	95.10
CNN [54]	85.98
Ours	95.29

In addition, Tables 6–8 show the precision, recall and f1-score of each model based on the dataset divided in the first way (test set of 828 images), respectively. As can be seen from Tables 6–8, our proposed model performed slightly better in the categories with less data. For example, from Table 7, it can be seen that our proposed model outperformed the IRv2-RA [11] and FixCaps [15] models by 83% and 33.3%, respectively, in the dermatofibroma (DF) skin disease category. Taken together, our proposed model outperformed all the other models in terms of the macro-average recall and macro-average f1-score. Meanwhile, in terms of macro-average precision, our proposed model was higher than most models, and it was only 0.79% and 0.59% lower than the IRv2-RA [11] and FixCaps [15] models, respectively.

Table 6. The precision of each model on the public dataset HAM10000 (test set of 828 images); the missing values of indicators are replaced by "-", and the unit is %.

Model	AKIEC	BCC	BKL	DF	MEL	NV	VASC	Macro-Average
ResNet50	58.33	69.70	75.41	66.67	56.52	94.44	75.00	70.87
EfficientNet_B4	100.00	75.00	82.81	66.67	66.67	96.57	100.00	83.96
DenseNet201	93.75	70.59	83.05	83.33	56.76	96.85	100.00	83.48
ConvNeXt_L	81.25	77.42	85.48	75.00	62.86	97.30	100.00	82.76
IRv2-RA [11]	100.00	88.00	72.00	100.00	67.00	97.00	100.00	89.14
FixCaps [15]	88.00	95.60	89.80	57.00	93.80	98.40	100.00	88.94
IM-CNN [22]	-	-	-	-	-	-	-	-
CNN [54]	-	-	-	-	-	-	-	84.00
Ours	87.50	85.71	96.30	85.71	65.22	98.03	100.00	88.35

Table 7. The recall of each model on the public dataset HAM10000 (test set of 828 images); the missing values of indicators are replaced by "-", and the unit is %.

Model	AKIEC	BCC	BKL	DF	MEL	NV	VASC	Macro-Average
ResNet50	30.43	88.46	69.70	33.33	38.24	97.44	90.00	63.94
EfficientNet_B4	52.17	92.31	80.30	100.00	58.82	97.74	100.00	83.05
DenseNet201	65.22	92.31	74.24	83.33	61.76	97.29	100.00	82.02
ConvNeXt_L	56.52	92.31	80.30	100.00	64.71	97.74	100.00	84.51
IRv2-RA [11]	52.00	88.00	83.00	17.00	65.00	98.00	100.00	71.86
FixCaps [15]	95.70	84.60	86.40	66.70	91.20	98.60	70.00	84.74
IM-CNN [22]	-	-	-	-	-	-	-	83.50
CNN [54]	-	-	-	-	-	-	-	86.00
Ours	91.30	92.31	78.79	100.00	88.24	97.44	100.00	92.58

Table 8. The f1-score of each model on the public dataset HAM10000 (test set of 828 images); the missing values of indicators are replaced by "-", and the unit is %.

Model	AKIEC	BCC	BKL	DF	MEL	NV	VASC	Macro-Average
ResNet50	40.00	77.97	72.44	44.44	45.62	95.92	81.82	65.46
EfficientNet_B4	68.57	82.76	81.54	80.00	62.50	97.15	100.00	81.79
DenseNet201	76.92	80.00	78.40	83.33	59.15	97.07	100.00	82.12
ConvNeXt_L	66.67	84.21	82.81	85.71	63.77	97.52	100.00	82.96
IRv2-RA [11]	69.00	88.00	77.00	29.00	66.00	98.00	100.00	75.29
FixCaps [15]	91.70	89.80	88.10	61.50	92.50	98.50	82.40	86.36
IM-CNN [22]	-	-	-	-	-	-	-	-
CNN [54]	-	-	-	-	-	-	-	85.98
Ours	89.36	88.89	86.67	92.31	75.00	97.73	100.00	89.99

Finally, Tables 9–12 show the accuracy, precision, recall and f1-score of each model on the dataset divided in the second way (test set of 2000 images), respectively. It can be observed from Tables 9–12 that our proposed model not only outperformed the models proposed by others in terms of accuracy but also outperformed the models proposed by others in terms of the macro-average precision, macro-average recall and macro-average f1-score. In particular, compared with the models proposed by others in terms of the macro-average recall and macro-average f1-score, our proposed model possessed the largest improvement of 18.91% and 14.75%, respectively. All in all, our proposed model not only possessed a good classification performance on our private dataset but also showed good classification performance on the public dataset HAM10000. In addition, compared with the other state-of-the-art models, it also achieved good results. This demonstrates that our proposed model possessed a good generalization ability at the same time. In order to facilitate the comparison of the classification performance of multiple models on multiple datasets, we performed a statistical analysis on the accuracy rates of the models ResNet50,

EfficientNet_B4, DenseNet201, ConvNeXt_L and our model on the above three datasets. The critical value calculated using the on-parametric Friedman test [55] was 0.0218. So, the test accuracy of these models showed a significant difference. Then, the post-hoc Nemenyi test [55] was used to further distinguish the model performance, where the calculated critical difference (CD) was 3.5215. The critical difference diagram is shown in Figure 9. According to Figure 9, it can be seen that our proposed model performed better overall.

Table 9. The accuracy of each model on the public dataset HAM10000 (test set of 2000 images); the unit is %.

Model	Accuracy
ResNet50	81.85
EfficientNet_B4	88.20
DenseNet201	87.75
ConvNeXt_L	88.40
Bayesian DenseNet169 [21]	83.59
MobileNetV2-LSTM [56]	85.34
EW-FCM and wide-ShuffleNet [32]	84.80
Shifted2-Nets [26]	83.60
Ours	90.85

Table 10. The precision of each model on the public dataset HAM10000 (test set of 2000 images); the missing values of indicators are replaced by "-", and the unit is %.

Model	AKIEC	BCC	BKL	DF	MEL	NV	VASC	Macro-Average
ResNet50	43.24	63.11	65.98	100.00	66.47	89.48	79.17	72.49
EfficientNet_B4	66.67	82.29	83.96	78.26	77.84	91.86	85.19	80.87
DenseNet201	64.38	84.62	82.70	94.44	70.81	92.55	79.31	81.26
ConvNeXt_L	56.00	81.90	76.96	82.35	76.67	94.37	92.00	80.04
Bayesian DenseNet169 [21]	-	-	-	-	-	-	-	-
MobileNetV2-LSTM [56]	-	-	-	-	-	-	-	-
EW-FCM and wide-ShuffleNet [32]	-	-	-	-	-	-	-	75.15
Shifted2-Nets [26]	-	-	-	-	-	-	-	76.00
Ours	64.77	93.26	83.57	85.71	81.45	95.37	82.14	83.75

Table 11. The recall of each model on the public dataset HAM10000 (test set of 2000 images); the missing values of indicators are replaced by "-", and the unit is %.

Model	AKIEC	BCC	BKL	DF	MEL	NV	VASC	Macro-Average
ResNet50	49.23	74.76	58.45	13.04	50.23	94.48	67.86	58.29
EfficientNet_B4	80.00	76.70	71.69	78.26	61.99	96.79	82.14	78.22
DenseNet201	72.31	85.44	69.86	73.91	66.97	95.38	82.14	78.00
ConvNeXt_L	64.62	83.50	76.26	60.87	72.85	95.08	82.14	76.47
Bayesian DenseNet169 [21]	-	-	-	-	-	-	-	-
MobileNetV2-LSTM [56]	-	-	-	-	-	-	-	-
EW-FCM and wide-ShuffleNet [32]	-	-	-	-	-	-	-	70.71
Shifted2-Nets [26]	-	-	-	-	-	-	-	64.90
Ours	87.69	80.58	81.28	78.26	81.45	95.30	82.14	83.81

Table 12. The f1-score of each model on the public dataset HAM10000 (test set of 2000 images); the missing values of indicators are replaced by "-", and the unit is %.

Model	AKIEC	BCC	BKL	DF	MEL	NV	VASC	Macro-Average
ResNet50	46.04	68.44	61.99	23.07	57.22	91.91	73.08	60.25
EfficientNet_B4	72.73	79.40	77.34	78.26	69.02	94.26	83.64	79.24
DenseNet201	68.11	85.03	75.74	82.92	68.84	93.94	80.70	79.33
ConvNeXt_L	60.00	82.69	76.61	70.00	74.71	94.72	86.79	77.93
Bayesian DenseNet169 [21]	-	-	-	-	-	-	-	-
MobileNetV2-LSTM [56]	-	-	-	-	-	-	-	-
EW-FCM and wide-ShuffleNet [32]	-	-	-	-	-	-	-	72.61
Shifted2-Nets [26]	-	-	-	-	-	-	-	68.70
Ours	74.51	86.46	82.41	81.82	81.45	95.33	82.14	83.45

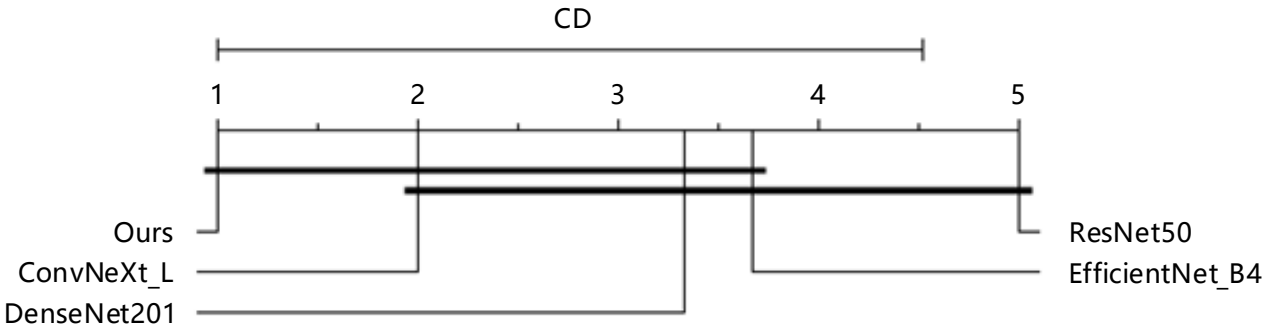


Figure 9. The critical difference diagram.

5. Discussion

Although our proposed model possessed good classification performance on the datasets with an extreme imbalance or a small number of samples, it was not flawless and still had limitations. For example, our proposed model consumed a lot of computing resources while training, and the training speed was also relatively slow. In addition, our proposed model recognized fewer types of skin diseases and requires training on more benchmark datasets in order to refine it. Therefore, in future work we will carry out a lightweight transformation of the proposed model in order to adapt it to different work scenarios. In addition, we will test our proposed model by using other benchmark datasets with different skin diseases.

6. Conclusions

In this paper, we proposed a convolutional neural network model for skin disease classification based on model fusion. We chose DenseNet201 and ConvNeXt_L as the backbone sub-classification models of our model fusion. In addition, on the core block of each sub-classification model, an attention module was introduced to assist the network in acquiring a region of interest in order to enhance the ability of the network model to extract image features. In addition, the features extracted by the shallow network could capture more details, and the features extracted by the deep network contained more abstract semantic information. Combining the characteristics of the two, a parallel strategy was adopted to fuse the features of the deep and shallow layers. Finally, through a series of works such as model pre-training, data augmentation and parameter fine-tuning, the classification performance of the proposed model was further improved.

On the private dataset, the proposed model achieved an accuracy of 96.49%, which was 4.42% and 3.66% higher than the two baseline models, respectively. On the public dataset, HAM10000, the accuracy and f1-scores of the proposed model were 95.29% and 89.99%, respectively, which also achieved good results compared to the other state-of-the-

art models. It was demonstrated that the proposed model possessed a good classification performance on the datasets with an extreme imbalance or a small number of samples as well as a good generalization ability.

Author Contributions: Conceptualization, M.W. and Q.W.; methodology, Q.W.; validation, H.J. and J.W.; formal analysis, M.W. and J.L.; writing—original draft preparation, Q.W.; writing—review and editing, J.L. and L.Z.; supervision, T.L.; project administration, T.L., L.Z., M.W., Q.W. and H.J. contribute equally to the work. All authors have read and agreed to the published version of the manuscript.

Funding: This publication emanated from research conducted with the financial support of the National Key Research and Development Program of China under grant no. 2017YFE0135700, the Tsinghua Precision Medicine Foundation under grant no. 2022TS003.

Institutional Review Board Statement: Ethical review and approval were waived for this study because the data used in this study only involved pictures of skin diseases, which do not involve ethics, and we did not involve experiments using animals.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: We analyzed a public dataset in this study. It is available at https://challenge.isic-archive.com/data/#2018, (access on 10 November 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Karimkhani, C.; Dellavalle, R.P.; Coffeng, L.E.; Flohr, C.; Hay, R.J.; Langan, S.M.; Nsoesie, E.O.; Ferrari, A.J.; Erskine, H.E.; Silverberg, J.I. Global skin disease morbidity and mortality: An update from the global burden of disease study 2013. *JAMA Dermatol.* **2017**, *153*, 406–412. [CrossRef] [PubMed]
2. Leiter, U.; Eigentler, T.; Garbe, C. Epidemiology of skin cancer. *Sunlight Vitam. D Ski. Cancer* **2014**, *810*, 120–140.
3. Baumann, B.C.; MacArthur, K.M.; Brewer, J.D.; Mendenhall, W.M.; Barker, C.A.; Etzkorn, J.R.; Jellinek, N.J.; Scott, J.F.; Gay, H.A.; Baumann, J.C. Management of primary skin cancer during a pandemic: Multidisciplinary recommendations. *Cancer* **2020**, *126*, 3900–3906. [CrossRef] [PubMed]
4. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J.A.; Van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [CrossRef] [PubMed]
5. Abdulrahman, A.A.; Rasheed, M.; Shihab, S. The Analytic of image processing smoothing spaces using wavelet. In Proceedings of the Ibn Al-Haitham International Conference for Pure and Applied Sciences (IHICPS), Baghdad, Iraq, 9–10 December 2020; p. 022118.
6. Rashid, T.; Mokji, M.M. Low-Resolution Image Classification of Cracked Concrete Surface Using Decision Tree Technique. In *Control, Instrumentation and Mechatronics: Theory and Practice*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 641–649.
7. Liu, W.; Wang, Z.; Liu, X.; Zeng, N.; Liu, Y.; Alsaadi, F.E. A survey of deep neural network architectures and their applications. *Neurocomputing* **2017**, *234*, 11–26. [CrossRef]
8. Pouyanfar, S.; Sadiq, S.; Yan, Y.; Tian, H.; Tao, Y.; Reyes, M.P.; Shyu, M.-L.; Chen, S.-C.; Iyengar, S.S. A survey on deep learning: Algorithms, techniques, and applications. *ACM Comput. Surv. (CSUR)* **2018**, *51*, 1–36. [CrossRef]
9. Ker, J.; Wang, L.; Rao, J.; Lim, T. Deep learning applications in medical image analysis. *IEEE Access* **2017**, *6*, 9375–9389. [CrossRef]
10. Anwar, S.M.; Majid, M.; Qayyum, A.; Awais, M.; Alnowami, M.; Khan, M.K. Medical image analysis using convolutional neural networks: A review. *J. Med. Syst.* **2018**, *42*, 226. [CrossRef]
11. Datta, S.K.; Shaikh, M.A.; Srihari, S.N.; Gao, M. Soft Attention Improves Skin Cancer Classification Performance. In *Interpretability of Machine Intelligence in Medical Image Computing, and Topological Data Analysis and Its Applications for Medical Data*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 13–23.
12. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.
13. Codella, N.C.; Gutman, D.; Celebi, M.E.; Helba, B.; Marchetti, M.A.; Dusza, S.W.; Kalloo, A.; Liopyris, K.; Mishra, N.; Kittler, H. 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 Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 168–172.
14. Tschandl, P.; Rosendahl, C.; Kittler, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data* **2018**, *5*, 180161. [CrossRef] [PubMed]
15. Lan, Z.; Cai, S.; He, X.; Wen, X. FixCaps: An Improved Capsules Network for Diagnosis of Skin Cancer. *IEEE Access* **2022**, *10*, 76261–76267. [CrossRef]
16. Sabour, S.; Frosst, N.; Hinton, G.E. Dynamic routing between capsules. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3856–3866.

17. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
18. 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, New Orleans, LA, USA, 18–24 June 2022; pp. 11976–11986.
19. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. Supplementary material for 'ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 13–19.
20. Yang, Z.; Zhu, L.; Wu, Y.; Yang, Y. Gated channel transformation for visual recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11794–11803.
21. Mobiny, A.; Singh, A.; Van Nguyen, H. Risk-aware machine learning classifier for skin lesion diagnosis. *J. Clin. Med.* **2019**, *8*, 1241. [CrossRef]
22. Wang, S.; Yin, Y.; Wang, D.; Wang, Y.; Jin, Y. Interpretability-based multimodal convolutional neural networks for skin lesion diagnosis. *IEEE Trans. Cybern.* **2021**, *52*, 12623–12637. [CrossRef] [PubMed]
23. Allugunti, V.R. A machine learning model for skin disease classification using convolution neural network. *Int. J. Comput. Program. Database Manag.* **2022**, *3*, 141–147.
24. Anand, V.; Gupta, S.; Koundal, D.; Nayak, S.R.; Nayak, J.; Vimal, S. Multi-class Skin Disease Classification Using Transfer Learning Model. *Int. J. Artif. Intell. Tools* **2022**, *31*, 2250029. [CrossRef]
25. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
26. Thurnhofer-Hemsi, K.; López-Rubio, E.; Domínguez, E.; Elizondo, D.A. Skin lesion classification by ensembles of deep convolutional networks and regularly spaced shifting. *IEEE Access* **2021**, *9*, 112193–112205. [CrossRef]
27. Karthik, R.; Vaichole, T.S.; Kulkarni, S.K.; Yadav, O.; Khan, F. Eff2Net: An efficient channel attention-based convolutional neural network for skin disease classification. *Biomed. Signal Process. Control* **2022**, *73*, 103406. [CrossRef]
28. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
29. Tan, M.; Le, Q. Efficientnetv2: Smaller models and faster training. In Proceedings of the International Conference on Machine Learning, Shenzhen, China, 26 February–1 March 2021; pp. 10096–10106.
30. Abayomi-Alli, O.O.; Damasevicius, R.; Misra, S.; Maskeliunas, R.; Abayomi-Alli, A. Malignant skin melanoma detection using image augmentation by oversamplingin nonlinear lower-dimensional embedding manifold. *Turk. J. Electr. Eng. Comput. Sci.* **2021**, *29*, 2600–2614. [CrossRef]
31. Mendonça, T.; Ferreira, P.M.; Marques, J.S.; Marcal, A.R.; Rozeira, J. PH 2-A dermoscopic image database for research and benchmarking. In Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, Japan, 3–7 July 2013; pp. 5437–5440.
32. Hoang, L.; Lee, S.-H.; Lee, E.-J.; Kwon, K.-R. Multiclass Skin Lesion Classification Using a Novel Lightweight Deep Learning Framework for Smart Healthcare. *Appl. Sci.* **2022**, *12*, 2677. [CrossRef]
33. Malibari, A.A.; Alzahrani, J.S.; Eltahir, M.M.; Malik, V.; Obayya, M.; Al Duhayyim, M.; Neto, A.V.L.; de Albuquerque, V.H.C. Optimal deep neural network-driven computer aided diagnosis model for skin cancer. *Comput. Electr. Eng.* **2022**, *103*, 108318. [CrossRef]
34. Nawaz, M.; Nazir, T.; Masood, M.; Ali, F.; Khan, M.A.; Tariq, U.; Sahar, N.; Damaševičius, R. Melanoma segmentation: A framework of improved DenseNet77 and UNET convolutional neural network. *Int. J. Imaging Syst. Technol.* **2022**, *32*, 2137–2153. [CrossRef]
35. Codella, N.; Rotemberg, V.; Tschandl, P.; Celebi, M.E.; Dusza, S.; Gutman, D.; Helba, B.; Kalloo, A.; Liopyris, K.; Marchetti, M. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). *arXiv* **2019**, arXiv:1902.03368.
36. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
37. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
38. Sagi, O.; Rokach, L. Ensemble learning: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1249. [CrossRef]
39. Zhou, Z.-H. Ensemble learning. In *Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 181–210.
40. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 10–17 October 2021; pp. 10012–10022.
41. Hendrycks, D.; Gimpel, K. Gaussian error linear units (gelus). *arXiv* **2016**, arXiv:1606.08415.
42. Ba, J.L.; Kiros, J.R.; Hinton, G.E. Layer normalization. *arXiv* **2016**, arXiv:1607.06450.
43. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
44. Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.

45. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]
46. DeVries, T.; Taylor, G.W. Improved regularization of convolutional neural networks with cutout. *arXiv* **2017**, arXiv:1708.04552.
47. Yun, S.; Han, D.; Oh, S.J.; Chun, S.; Choe, J.; Yoo, Y. Cutmix: Regularization strategy to train strong classifiers with localizable features. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6023–6032.
48. Hendrycks, D.; Mu, N.; Cubuk, E.D.; Zoph, B.; Gilmer, J.; Lakshminarayanan, B. Augmix: A simple data processing method to improve robustness and uncertainty. *arXiv* **2019**, arXiv:1912.02781.
49. Zhong, Z.; Zheng, L.; Kang, G.; Li, S.; Yang, Y. Random erasing data augmentation. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 13001–13008.
50. Bishop, C.M.; Nasrabadi, N.M. *Pattern Recognition and Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2006; Volume 4.
51. Olson, D.L.; Delen, D. *Advanced Data Mining Techniques*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2008.
52. Opitz, J.; Burst, S. Macro f1 and macro f1. *arXiv* **2019**, arXiv:1911.03347.
53. Bottou, L. Stochastic gradient descent tricks. In *Neural Networks: Tricks of the Trade*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 421–436.
54. Alwakid, G.; Gouda, W.; Humayun, M.; Sama, N.U. Melanoma Detection Using Deep Learning-Based Classifications. *Healthcare* **2022**, *10*, 2481. [CrossRef] [PubMed]
55. Demšar, J. Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* **2006**, *7*, 1–30.
56. Srinivasu, P.N.; SivaSai, J.G.; Ijaz, M.F.; Bhoi, A.K.; Kim, W.; Kang, J.J. Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. *Sensors* **2021**, *21*, 2852. [CrossRef] [PubMed]