Multiscale ensemble of convolutional neural networks for skin lesion classification

Yi-Peng Liu | Ziming Wang | Zhanqing Li | Jing Li | Ting Li | Peng Chen | Ronghua Liang

1 College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China
2 College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China
3 Cancer Institute of Integrative Medicine, Tongde Hospital of Zhejiang Province, Hangzhou 310012, China
4 Department of Pharmacy, Hangzhou Xixi Hospital Affiliated to Zhejiang Chinese Medical University, Hangzhou 310023, China

Correspondence
Jing Li, Cancer Institute of Integrative Medicine, Tongde Hospital of Zhejiang Province, Hangzhou 310012, China.
Email: lijing.td@aliyun.com

Funding information
Natural Science Foundation of China, Grant/Award Number: 62076220; Zhejiang Provincial Chinese Medicine Science and Technology Project, Grant/Award Number: 2021ZB059; Fundamental Research Funds for the Provincial Universities of Zhejiang, Grant/Award Number: RF-C2019001

Abstract
Early detection and treatment of skin cancer can considerably reduce the patient mortality rates. Convolutional neural network (CNN) has been widely applied in the field of computer aided diagnosis. However, for skin lesions, the inconsistent size of lesion regions in dermoscopic images hinders the convolutional neural network precise discrimination. To solve this problem, multiscale ensemble of convolutional neural networks called MECNN is proposed, which involves three branches with different lesion scales as the model input. The first branch locates the lesion region outline by identifying the largest local response point. Then, MECNN reduces the search area of the lesion region and divides the outline into two scales used as the input for the other two branches. A global loss function is defined to control the learning objectives of the three branches and MECNN fuses the branches output as the final classification result. The proposed model is evaluated on the public HAM10000 dataset and achieves a higher classification accuracy than the comparative state-of-the-art methods.

1 INTRODUCTION

Skin cancer is one of the most widespread malignant tumours caused by exposure to UV rays from the sun or chemical stimulation etc. that damage the DNA of the skin cells [1]. Melanoma is the most deadly skin cancer, but the 5-year survival rate of early stage melanoma is more than 95%. Therefore, early detection and diagnosis of skin cancer is highly significant and will reduce the mortality rate [2].

Skin cancer is examined visually by clinical experts starting from preliminary screening with the following dermoscopic assessment [3]. Dermoscopy is a non-invasive technology, which can capture high-resolution images of the skin and enable dermatologists to detect invisible features. This technology is particularly important in the diagnosis of melanoma. However, visual diagnosis is very time-consuming and subjective. It is difficult for dermatologists to identify malignant and benign skin lesions because of the visual similarity [4]. Therefore, computer-aided diagnosis (CAD) system for skin lesion recognition has been naturally proved to be a contributing assessment tool that could reduce inter-observer variability and address the limited availability of trained experts [5].

Currently, convolutional neural network (CNN) exhibits a notably high performance in many computer vision applications, including image classification [6, 7], image segmentation [8, 9] and target detection [10, 11]. Many researchers have employed CNN to classify skin lesions, but are still plagued by the challenging issues [2, 12, 13]: (1) the high degree of
intra-class variation and inter-class similarity; (2) various artefacts including hair, colour illumination etc.; (3) the poor generalization ability of the CNN; (4) highly imbalanced lesion classes.

To relieve these problems, we propose a skin lesion classification method called MECNN based on multiscale ensemble of CNNs. The proposed framework includes three branches (one main branch and two auxiliary branches), all of which could be embedded by the common network structures, such as ResNet [6] and DenseNet [7]. We reduce the dimensions of the features extracted from the main branch to obtain the attention map, identify the position of the maximum local response point in the attention map and map this point to the input image to obtain the window cropping centre. Subsequently, we crop the input image with two different window sizes to obtain two scales images A global loss function is defined to control the learning objectives of the three branches, owing to which, the prediction results of the two auxiliary branches as similar as possible to those of the main branch. The main contributions are summarized as follows:

1. We propose a skin lesion classification method MECNN based on multiscale ensemble of convolutional neural networks, which achieves the better results compared with state-of-art methods on the public dataset.
2. The proposed framework can be assembled with common CNN structures and substantially improves their classification performances.
3. A global loss function is designed to control the learning objectives of multiscale network branches and instructive to reduce intra-class differences.
4. The proposed multiscale ensemble can accurately extract the salient lesion region in dermatologic images without extra segmentation labels data and facilitate to relieve the noise interference from background around the lesion region.

The remaining paper is organized as follows: Section 2 presents a review of related works and summarizes the existing research on the classification of skin lesions. Section 3 describes the proposed method. Section 4 discusses the experimental results and the related evaluation. Finally, the proposed approach is concluded in Section 5.

2 RELATED WORKS

2.1 Skin lesion classification

In recent years, deep learning has been widely used in the field of skin lesion classification. Kawahara et al. [14] proposed a multi-channel CNN where multiresolution images were provided as an input, and the framework was optimized by an auxiliary loss function. Esteva et al. [15] trained InceptionV3 with 129,450 images to diagnose the most common and fatal skin cancers, and the diagnostic level of this method was noted to be similar to that of dermatologists. Ge et al. [16] trained a CNN model with clinical images and demonstrated the effectiveness of multimodal learning in skin lesion classification. Zhang et al. [17] proposed attention residual learning based CNN for skin lesion classification. This method improved the last layer of ResNet and enhanced the model's ability to discriminate dermatologic images through a novel attention mechanism. Menegola et al. [18] used six public datasets to pretrain CNNs to boost the performance of skin lesion classification. Chaturvedi et al. [19] trained a pretrained MobileNet on the HAM10000 dataset [20]. This model outperformed the existing models and exhibited the higher speed and lower parameters size. Zhang et al. [21] optimized a collaborative deep learning method in terms of the classification and collaborative errors, effective in many medical image classification tasks. Ratul et al. [22] employed dilated convolution and reported that the highest performance compared the extended InceptionV3.

2.2 Segmentation based lesion classification

To focus on the local lesion region of dermatologic images, many of the existing methods first locate the skin lesion region and subsequently classify it [5, 13]. Diaz et al. [23] used a structure segmentation model to enable further feature representation. The diagnostic model used all the features to predict the final skin lesion type. Yu et al. [24] used the segmentation-classification union model to increase the accuracy. Specifically, the authors used the segmentation mask to locate the skin lesion region and classified the cropped skin lesion region image. Jia et al. [25] proposed a two stage framework with only one network to classify the dermatologic images. A convolution network was trained and subsequently the input region with the maximum activation value was cut out, and the network was retrained to output the final probability. Alom et al. [26] combined the feature mapping from lower and higher layers and achieve excellent results with the same or less network parameters in both segmentation and classification tasks.

2.3 Ensemble learning based classification

To enhance the classification accuracy of skin lesions, many studies used ensemble learning to fuse the diagnosis results of multiple methods [27]. Matsunaga et al. [28] proposed a deep neural network ensemble method to classify melanoma, seborrheic keratosis and nevus. First, a classifier was generated to classify melanoma and other lesions. Second, another classifier was generated to classify seborrheic keratosis and other skin injuries. Finally, the model fused the output classification probability. Mahbod et al. [29] proposed a CNN integration scheme, which used the AlexNet, VGG16 and ResNet-18 models to extract features. Subsequently, the features were used to train different support vector machine classifiers, and finally the classification vectors were fused to provide the final classification results. Harangi et al. [30] combined the classification layer outputs of GoogLeNet, AlexNet, ResNet and VGGNet, and the combined model attained a high classification accuracy than other fusion techniques. Hagerty et al. [31] recommended
3 MATERIALS AND METHODS

3.1 Data sets

In this study, the HAM10000 dataset is used, which contains 10,015 images with labels pertaining to seven categories: (a) Melanoma (MEL), (b) melanocytic nevus (NV), (c) basal cell carcinoma (BCC), (d) actinic keratoses and intraepithelial carcinoma (AKIEC), (e) benign keratosis (BKL), (f) dermatofibroma (DF), and (g) vascular lesions (VASC). Table 1 indicates that the dataset is unevenly distributed, with the maximum and minimum number of images being 6705 and 115 in different classes, respectively.

3.2 MECNN

To solve the problem of the deteriorated CNN performance, caused by the inconsistent size of the lesion regions in dermatoscope images, we propose a skin lesion classification method based on multiscale ensemble CNNs called MECNN. As shown in Figure 1. MECNN includes three branches (one main branch and two auxiliary branches), and the feature extraction modules of each branch share the same structure. The feature extraction module can be substituted by common network structures, such as ResNet, DenseNet etc.

Specifically, an input image passes through the main branch’s feature extractor module to obtain feature maps. These feature maps are then passed through a classifier to obtain the classification output $P_i$ and used to generate an attention map. In this attention map, we can identify the position point that having maximum local feature response and map this point to input image to locate the cropping centre. Two images Crop1 and Crop2 with different scales are obtained by cropping the input image with the cropping centre and two preset cropping window sizes. These two cropped images are later used as the inputs of the two auxiliary branches.

Before feeding to auxiliary branches, Crop1 and Crop2 will first be upsampled into the size of the input image. Then, they are sent to the feature extractor module and classifier like the main branch to obtain $P_i^2$ and $P_i^3$. In the training stage, we expect that $P_i^2$ and $P_i^3$ obtained in auxiliary branches can improve the lesion location ability of main branch. Therefore, a global loss function is designed to control the learning objectives of the three branches, owing to which, the predicted results of the two auxiliary branches are similar to those of the main branch. The same network structure is used in the test stage, and the results can be combined with the multiscale image information, which is reflected in the image cropping process. The multiscale information provides an attention based data augmentation way, and the cropping window is instructive to relieve the noise interference from background around the lesion region. The multiscale pairwise ranking loss functions facilitate to reduce intra-class differences. The pseudo code in Table 2 illustrates our method further.

A detailed description of the method is presented as follows.

3.2.1 Data augmentation

We resize the input image with $224 \times 224$. The images of each category are randomly divided into 5 equal groups, of which 4 groups are used for training and the last group is used for testing. Because of unbalanced distribution of the training dataset, we augment the number for each image category to the similar number. The specific mechanisms include rotation or contrast, brightness or saturation adjustment, and thus the total training set capacity becomes 59,500 images.

3.2.2 Cropping scheme

The proposed method aims at solving the problem of the scales non-uniformity of the lesion regions in dermatologic images; therefore, it is crucial to suitably extract the lesion regions. We cut out the lesion region by setting cropping boxes of different sizes to obtain cropped images. The obtained images are upsampled into same size of original image, thereby enlarging the lesion region. The specific operation is shown in Figure 2. First, we seek out the max local feature response point from feature maps Figure 2(A). Then, we map the point to the input image Figure 2(C) as the cropping centre. Finally, we crop the input image with the cropping centre and preset window size. The

| MEL | NV | BCC | AKIEC | BKL | DF | VASC |
|-----|----|-----|-------|-----|----|------|
| 1113| 6705| 514 | 327   | 1099| 115| 142  |
FIGURE 1  Network structure diagram overview. The features of the input image a1 are extracted to obtain the attention map, and then cropped on the attention map to obtain Crop1 and Crop2 with different scales, which are used to further extract features. $P_r^1, P_r^2, P_r^3$ represent the execution degrees of the output. $L_s$ and $L_{pair}$ represent the classification and pairwise ranking loss functions, respectively.

TABLE 2  The pseudo code of MECNN. $S$ is the total number of training steps, $N$ is the number of auxiliary branches, $h_n, w_n$ are the cropping windows size of an auxiliary branch.

**Forward propagation**

for step $s = 1, S$ do:
  get input image $x$
  get max feature response point with Equation (1)
  get crop centre with Equation (2)
  for $n = 1, N$ do:
    crop $x$ with the cropping centre, $h_n, w_n$
    obtain $P_r^1, P_r^2, P_r^3$
  calculate loss with Equation (3)

process of finding the max local feature response point can be formulated as

$$\left( x_{mr}, y_{mr} \right) = \arg \max_{x, y} f_{3 \times 3}(M_{x,y})$$  \hspace{1cm} (1)

where $M_{x,y}$ is the output feature maps before FC layer of the main branch in Figure 1, and $(x, y)$ is the space coordinate of $M$. $f_{3 \times 3}$ is a convolution operation with kernel size of $3 \times 3$ and the fixed kernel weights of 1. After finding the max local feature response point $(x_{mr}, y_{mr})$, we map this point to input image by

$$\left( x_c, y_c \right) = \left( x_{mr}, y_{mr} \right) \times k$$  \hspace{1cm} (2)

where $(x_c, y_c)$ is the cropping centre in input image, $k$ is the downsampling times in feature extraction module. Note that if the cropping box exceeds input image, the vertex closest to the orange point in A is selected, and cropping is performed considering this vertex as that of the cropping box.

3.2.3  Loss function

A global loss function is designed, motivated by Fu et al. [34]. The total loss function consists of two parts, namely, classification and pairwise ranking loss functions. The loss function is defined as

$$L = \sum_{i=1}^{3} L_s + \sum_{i=2}^{3} \left\{ L_{pair}(P_r^1, P_r^i) \right\}$$  \hspace{1cm} (3)
where \( s \) represents the scale, and \( L_s \) represents the classification loss. By reducing the classification loss in the training process, the parameters of the convolution and classification layers in Figure 1 are optimized to ensure that each branch has a sufficient recognition ability. Moreover, \( p_{r_1}, p_{r_2} \) and \( p_{r_3} \) in the comparison loss function \( L_{\text{pair}} \) represent the prediction probabilities of the correct labels in the outputs of the three branches shown in Figure 2. The pairwise loss functions are defined as

\[
L_{\text{pair}}(p_{r_1'}, p_{r_1}) = \max\{p_{r_1'} + \text{margin} - p_{r_1}, 0\}
\]

(4)

This constraint ensures that \( p_{r_1'} > p_{r_1} + \text{margin} \) during the training stage, and the hyperparameter margin is set to 0.05 in all the experiments. In the training process, as shown in Figure 1, the prediction results of the lower two branches are as similar as possible to those of the upper branch, and the final classification results are obtained by comprehensively considering the outputs of the three branches.

4 | EXPERIMENTAL RESULTS

As the parameter setup for all the experiments, the initial learning rate is set to 0.001, and an attenuation of 0.1 times is implemented for every 7 epochs. The SGD optimizer is used to optimize the training process. We find that the margin in Equation (4) is robust to optimization, thus we empirically set margin as 0.05. All the experiments are performed on the Ubuntu system with a 1080Ti GPU and PyTorch framework.

4.1 | Evaluation indices

The evaluation indices of the experiments include the precision, recall and F1-score. TP, TN, FP and FN, represent the true positive, true negative, false positive and false negative, respectively.

\[
\text{Precision} = \frac{TP}{TP + FP}
\]

(5)

\[
\text{Recall} = \frac{TP}{(TP + FN)}
\]

(6)

\[
\text{F1 - Score} = 2 \times \frac{\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}}{1}
\]

(7)

4.2 | Compared with recent skin lesion classification methods

We use different network structures (ResNet50, ResNet101, DenseNet121, DenseNet161) as feature extraction modules. The evaluation index is calculated by micro-average and macro-average methods. Micro-average considers all categories at one time to calculate the accuracy of category prediction, while macro-average considers each category separately, and obtain the final accuracy of the test set by arithmetic average. Because both micro-average and macro-average are used for comparison in the field of skin lesions classification, we present the results of both average methods as listed in Figure 3 and Table 3. Considering the precision, recall and F1-score of each model, as listed in Table 3, the largest values in the training process of 50 epochs are extracted. Moreover, the proposed method is compared with other ensemble methods, with [en] denoting the ensemble method.

Even when different network structures are used as the feature extraction modules, the proposed framework could achieve higher performance, which demonstrates that the proposed multiscale deep integration model is not susceptible to the feature extraction modules. In the model with the best performance, ResNet50 is used as the ensemble model of the feature extraction module. The corresponding macro precision, macro recall and macro F1-score are 0.8941, 0.8769 and 0.8723, respectively. Among the other skin lesion classification methods, Gessert’s method exhibits the highest performance.

4.3 | Lesion region cropping results

Because of the inconsistent size of the lesion regions in dermatologic images, especially the extremely small lesion regions...
Table 3: Comparison of MECNN framework and recent studies with macro-average calculation. We use different basic networks as the feature extraction modules. The maximum values of the performance indices are boldfaced.

| Methods            | Precision | Recall  | F1-score |
|--------------------|-----------|---------|----------|
| ResNet50           | 0.8258    | 0.8143  | 0.8142   |
| ResNet50-Our       | 0.8941    | 0.8769  | 0.8723   |
| ResNet101          | 0.8347    | 0.8412  | 0.8323   |
| ResNet101-Our      | 0.8936    | 0.8674  | 0.8628   |
| DenseNet121        | 0.8483    | 0.8207  | 0.8237   |
| DenseNet121-Our    | 0.8939    | 0.8663  | 0.8627   |
| DenseNet161        | 0.8343    | 0.8229  | 0.8218   |
| DenseNet161-Our    | 0.8844    | 0.8589  | 0.8624   |
| Esteva et al. [15] | 0.7520    | 0.7108  | 0.7123   |
| Ratul et al. [22]  | 0.8800    | 0.7500  | 0.8100   |
| Al-masni et al. [5] | –        | 0.8100  | 0.8128   |
| Chaturvedi et al. [19]| 0.8900  | 0.8300  | 0.8300   |
| Zhang et al. [17]  | 0.8599    | 0.8284  | 0.8351   |
| Damian et al. [36] | 0.8874    | 0.7923  | 0.8371   |
| Gessert et al. [4] | –         | 0.7570  | 0.8610   |
| [en]Carcagni et al. [33]| 0.8800  | 0.7600  | 0.8200   |
| [en]Heller et al. [35]| 0.8430  | 0.8730  | 0.8577   |

The problem of reducing the interference of hair in skin dermatologic images has challenged researchers for a long time. The hair interference directly affects the results. Therefore, we crop the image with hair interference and observe whether the result is affected by the hair, which leads to the inaccurate extraction of the lesion region. It is noted that the cropping frame is barely affected by the hair in Figure 5.

4.4 Confusion matrix analysis

To further analyse the effectiveness of the method, we study the confusion matrix. Zhang et al. [17] also used ResNet50 as the feature extraction module like our method, so we compare the result with the method [17]. The experimental results are shown in Figure 6.
The proposed method exhibits the higher accuracy for each category, which shows that the proposed method is effective for all the skin lesion categories. For the NV columns, it is noted that the condition is less than the results of [17] when the proposed method judges the other classes as the NV classes. This finding indicates that the proposed method can accurately locate the lesion region by cropping out the features at different scales after extracting the features, and thus the features of each class can be extracted more effectively when re-extracting the features. Moreover, the results obtained at different scales can be combined in the final classification, thereby improving the accuracy for each class.

4.5 Ablation study

Here, we discuss the effectiveness of the cropping process and the pairwise ranking loss function in our methods.

4.5.1 The cropping process

The Table 4 shows the performance when choosing different windows sizes and choosing different number of sub-branches. According to the result of using only one sub-branch, the windows size smaller than 140 lead to the worse performance that will crop a limited area from lesion region. Based on this result, we choose two window size 168 and 196 to cover the whole range from 140 to 224 (the original image size) and get our best result.

Besides, we show the influence of the cropping centre selection in Figure 7, the green line represents the F1-score result that in our method we use fixed point (input image centre) as the cropping centre. Results show that the attention map based centre selection is beneficial to find the lesion region and improve the F1-score performance with 2–5%. The performance based on the fixed cropping centre strongly depends on the location
**Figure 6** Confusion matrix results: (a) Zhang et al. [17]; (b) the proposed method

**Figure 7** The influence of the cropping centre selection. The green line represents F1-score using fixed centre point and the red line represents F1-score using the maximum feature response point as the cropping centre.
and size of the lesion area. However, our method based on the local maximum response point of the attention map can dynamically adjust the cropping window position according to lesion location.

4.5.2 The pairwise ranking loss function

In Table 5, we list the prediction probability of our model based on ResNet50 and two branches for 10 different dermatologic images. All the model prediction probabilities are notably higher than the main branch, which enhances the model prediction confidence beneficial from the pairwise ranking loss.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we proposed a multiscale deep ensemble model for skin lesion classification. The network can accurately crop out the lesion regions in dermatologic images for further feature extraction and classify the results by combining the features obtained at different scales. The proposed method outperforms the other methods on the dermatologic dataset HAM10000. Future work will be focused on accurately locating the skin lesion regions by dynamically adjusting the size of the cropping box.

ACKNOWLEDGMENTS

This work was financially supported in part by Natural Science Foundation of China (62076220), Zhejiang Provincial Chinese Medicine Science and Technology Project (2021ZB059), the Fundamental Research Funds for the Provincial Universities of Zhejiang (RF-C2019001). The authors would like to thank Wangyang Zuo for his kind help on the paper revision.

ORCID

Yi-Peng Liu https://orcid.org/0000-0001-8658-5764
Peng Chen https://orcid.org/0000-0001-6122-0574

REFERENCES

1. Albahar, M.A.: Skin lesion classification using convolutional neural network with novel regularizer. IEEE Access 7, 38306–38313 (2019)
2. Tang, P., et al.: GP-CNN-DTEL: Global-part CNN model with data-transformed ensemble learning for skin lesion classification. IEEE J. Biomed. Health. Inf. 24(10), 2870–2882 (2020)
3. Iqbal, I., et al.: Automated multi-class classification of skin lesions through deep convolutional neural network with dermoscopic images. Comput. Med. Imaging Graphics 88, 101843 (2021)
4. Gessert, N., et al. Skin lesion classification using CNNs with patch-based attention and diagnosis-guided loss weighting. IEEE Trans. Biomed. Eng. 67(2), 495–503 (2020)
5. Al-masni, M.A., Kim, D.H., Kim, T.S.: Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification. Comput. Methods Programs Biomed. 190, 105351 (2020)
6. He, K., et al.: Deep residual learning for image recognition. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, pp. 770–778 (2016)
7. Huang, G., et al.: Densely connected convolutional networks. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, pp. 2261–2269 (2017)
8. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, pp. 3431–3440 (2015)
9. Chen, L.C., et al.: DeepLab: Semantic image segmentation with deep convolutional nets, Atrous convolution, and fully connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. 40(4), 834–848 (2017)
10. He, K., et al.: Mask R-CNN. In: Proceedings of IEEE International Conference on Computer Vision (ICCV). Venice, Italy, pp. 2980–2988 (2017)
11. Lin, T.Y., et al: Focal loss for dense object detection. In: Proceedings of IEEE International Conference on Computer Vision (ICCV). Venice, Italy, pp. 2999–3007 (2017)
12. Wu, J., et al. Skin lesion classification using densely connected convolutional networks with attention residual learning. Sensors 20, 7080 (2020)
13. Wu, H., et al.: Automated skin lesion segmentation via an adaptive dual attention module. IEEE Trans. Med. Imaging 40(1), 357–370 (2021)
14. Kawahara, J., Hamarneh, G.: Multi-resolution-tract CNN with hybrid pretrained and skin-lesion trained layers. In: International Workshop on Machine Learning in Medical Imaging (MLMI). Athens, Greece, pp. 164–171 (2016)
15. Esteva, A., et al.: Dermatologist-level classification of skin cancer with deep neural networks. Nature 542(7639), 115–118 (2017)
16. Ge, Z., et al.: Skin lesion recognition using deep saliency features and multimodal learning of dermoscopy and clinical images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Quebec, Canada, pp. 250–258 (2017)
17. Zhang, J., et al.: Attention residual learning for skin lesion classification. IEEE Trans. Med. Imaging 38(9), 2092–2103 (2019)
18. Menegola, A., et al.: RECOD Titans at ISIC challenge 2017. arXiv:1703.04819 (2017)
19. Chaturvedi, S.S., Gupta, K., Prasad, P.S.: Skin lesion analyser: An efficient seven-way multi-class skin cancer classification using mobiLeNet. In: International Conference on Advanced Machine Learning Technologies and Applications (AMLTA); Jaipur, India, pp. 165–176 (2020)
20. Tschantl, P., Rosendahl, C., Kittler, H.: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci. Data 5, 180161 (2018)
21. Zhang, J., et al: Medical image classification using synergetic deep learning. Med. Image Anal. 54, 10–19 (2019)
22. Ratul, A.R., et al.: Skin lesions classification using deep learning based on dilated convolution. bioRxiv:860700 (2019)
23. Diaz, L.G.: Incorporating the knowledge of dermatologists to convolutional neural networks for the diagnosis of skin lesions. arXiv:1703.01976 (2017)
24. Yu, L., et al.: Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Trans. Med. Imaging 36(4), 994–1004 (2017)
25. Jia, X., Shen, L.: Skin lesion classification using class activation map. arXiv:1703.01053 (2017)
26. Alom, M.Z., et al.: Skin cancer segmentation and classification with NABLA-N and Inception recurrent residual convolutional networks. arXiv:1904.11126 (2019)
27. Gessert, N., et al. Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data. MethodsX 7, 100864 (2020)
28. Matsunaga, K., et al.: Image classification of melanoma, nevus and seborrheic keratosis by deep neural network ensemble. arXiv:1703.03108 (2017)
29. Mahbod, A., et al.: Fusing fine-tuned deep features for skin lesion classification. Comput. Med. Imaging Graphics 71, 19–29 (2019)
30. Harangi, B.: Skin lesion classification with ensembles of deep convolutional neural networks. J. Biomed. Inf. 86, 35–52 (2018)
31. Hagerty, J.R., et al.: Deep learning and handcrafted method fusion: Higher diagnostic accuracy for melanoma dermoscopy images. IEEE J. Biomed. Health. Inf. 23(4), 1385–1391 (2019)
32. Bi, L., et al.: Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks. arXiv:1703.04197 (2017)
33. Carcagni, P., et al.: Classification of skin lesions by combining multilevel learnings in a DenseNet architecture. In: Proceedings of International Conference on Image Analysis and Processing (ICIAP). Trento, Italy, pp. 335–344 (2019)
34. Fu, J., Zheng, H., Mei, T.: Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, pp. 4438–4446 (2017)
35. Heller, N., et al.: Computer aided diagnosis of skin lesions from morphological features. Semantic Scholar, Corpus ID: 52108303 (2018). https://www.semanticscholar.org/paper/Computer-Aided-Diagnosis-of-Skin-Lesions-from-Heller-Bussmann/91eeb0ca7d0d3485ee000055f92741140266aa58
36. Almaraz-Damian, J.A., et al.: Melanoma and nevus skin lesion classification using handcraft and deep learning feature fusion via mutual information measures. Entropy 22, 484 (2020)

How to cite this article: Liu, Y.-P., et al.: IET Image Process. 15, 2309–2318 (2021).
https://doi.org/10.1049/ipt2.12214