Clinical-Inspired Network for Skin Lesion Recognition

Zihao Liu[1,2], Ruiqin Xiong[1], and Tingting Jiang[1]

[1] NELVT, Department of Computer Science, Peking University, Beijing, China
{lzh19961031,ttjiang}@pku.edu.cn
[2] Advanced Institute of Information Technology, Peking University, Hangzhou, China

Abstract. Automated skin lesion recognition methods are useful for improving the diagnostic accuracy in dermoscopy images. However, several challenges delayed the pace of the development of these methods, including limited amount of data, a lack of ability to focus on the lesion area, poor performance for distinguishing between visually-similar categories of diseases and an imbalance between different classes of training data. During practical learning and diagnosis process, doctors conduct certain strategies to tackle these challenges. Thus, it's really appealing to involve these strategies in automated skin lesion recognition method, which could be promising for a better performance. Inspired by this, we propose a new Clinical-Inspired Network (CIN) to simulate the subjective learning and diagnostic process of doctors. To mimic the diagnostic process, we design three modules, including a lesion area attention module to crop the images, a feature extraction module to extract image features and a lesion feature attention module to focus on the important lesion parts and mine the correlation between different lesion parts. To simulate the learning process, we introduce a distinguish module. The CIN is extensively tested on ISIC2016 and 2017 challenge datasets and achieves state-of-the-art performance, which demonstrates its advantages.

Keywords: Clinical-inspired · Skin lesion recognition · Neural networks

1 Introduction

Skin disease is one of the most common diseases in the world, and the number of deaths by skin diseases has increased in recent years, which aroused public attention [15, 14]. Dermoscopy [10, 2, 16] is an essential means of improving the recognition accuracy of skin diseases for doctors. A lot of efforts are dedicated to automating the classification of dermoscopy images since the manual inspection suffers from subjective bias.

Recently, many CNN-based methods come into being and achieve significant progress in classifying skin diseases [6, 19, 5, 22, 9, 20, 21, 17, 23]. One category of these methods is from the perspective of data. A synergic loss is introduced in [22, 9] to tackle the inter-class similarity and intra-class variation problem in skin lesion data. To solve the imbalance problem among different classes, the authors in [19] present to dynamically sample paired data and learn domain-invariant features. The second category deliberates the design of network architecture. A deep two-stage network is proposed in [20] to acquire richer and more discriminative features for more accurate recognition. It segments the images first, then classifies them. To enhance the network's ability for discriminative representation, the attention mechanism including spatial and channel attention is

integrated in [23]. A joint learning method based on both handcrafted and deep-learning features is proposed in [17] in order to construct the hybrid-prior feature representation for a better result. To take advantage of the multiple features from single image and build a global image representation, authors in [21] adopt Fisher Vector (FV) encoding to aggregate the orderless visual statistic features.

Although these methods have achieved good performance, the perspective of designing the algorithm in existing methods is similar to which for tackling natural image classification tasks, which is not fully driven by medical perspective and knowledge and ignores the medical expertise of skin disease. However, the classification task in dermoscopy images should be treated differently because it requires a lot of professional medical knowledge. Thus, having medical professionals' knowledge as much as possible involved in computer-aided methods becomes more and more appealing and significant. As a result, designing the network in a way that allows it to mimic all the processes of professionals, could let it learn and classify skin diseases more effectively.

Motivated by this, medical doctors' knowledge learning process and diagnostic process have been adopted and migrated by our method. On the side of doctors' diagnostic process, we find three steps for doctors to diagnose skin diseases. When a doctor looks at a dermoscopy image, first, in order to focus on the lesion area, he zooms in and only focuses on the lesion area, which can be called **zoom step**. Then, he observes and examines the feature of the lesion area, including shape, structure and color, which are all classical features for skin diseases [18]. This is called **observe step**. At last, the doctors will focus on the important lesion parts and mine the correlation between lesion parts before making a final diagnosis for a comprehensive judgment. This can be called **compare step**. On the side of doctors' learning process, it's really difficult to distinguish between different classes since the visual differences between images from the same kind of disease can be significantly greater than from the other kind. In order to solve this problem, doctors are devoted to comparing images which may visually look similar and decide whether they're from the same class. This is **distinguish strategy**.

In regard of all the information, we propose the Clinical-Inspired Network(CIN). **To simulate the diagnostic process, we introduce three modules.** Firstly, a lesion area attention module is adopted to mimic the zoom step, which enables the network to focus on the lesion area. To simulate the observe step, a convolutional neural network is used to extract features of input images, which forms the feature extraction module. For the compare step, the lesion feature attention module is introduced to focus on the important lesion parts and mine the correlation between different lesion parts. **To mimic the distinguish strategy in the learning process**, a distinguish module is proposed to increase the discrimination ability among different classes.

The main contributions can be summarized as follows: (1) We propose a Clinical-Inspired Network(CIN) on the perspective of simulating doctor's learning and diagnostic processes. (2) Three new modules are designed in CIN, including a lesion area attention module, a lesion feature attention module and a distinguish module, which is fully inspired by the clinical process of doctors. To sum up, our method is different from existing methods both in perspective and architecture. Our perspective is especially unique and important for medical problems since few former methods are driven by medical perspective and knowledge and systematically consider all the processes of

doctors. The experimental results show that our method performs the best on both ISIC 2016 and 2017 challenge datasets.

2 Proposed Method

The whole network architecture of CIN is illustrated in Fig. 1. It consists of four modules. The first module is the Lesion Area Attention Module. It takes two original dermoscopy images as input, generates the lesion area attention map for each image and takes the smallest rectangle which includes the lesion area from the attention map to crop out the images. The following Feature Extraction Module extracts the features of the two cropped images. Next, The feature of each image is fed into the Lesion Feature Attention Module, which includes a spatial attention block and a correlation attention block. After that, the Distinguish Module takes the output features of Lesion Feature Attention Module from two branches as inputs, compares them and outputs a score reflecting whether two images are from the same class. Meanwhile, the two features are fed into two fully connected layers respectively, in order to do the classification task.

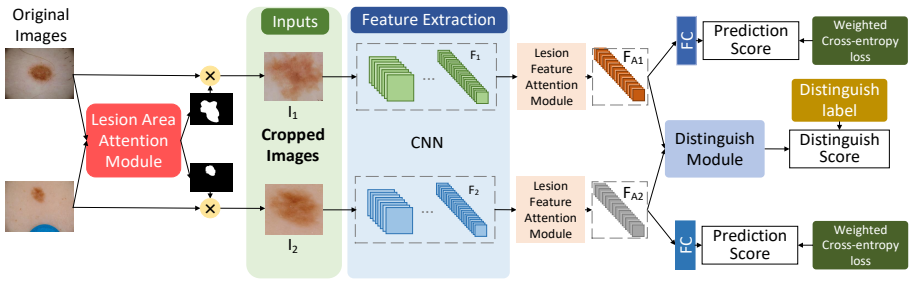

Fig. 1. An overview of the CIN.

Lesion Area Attention Module This module is adopted to mimic the **zoom step** of diagnostic process for doctors. As illustrated in Fig. 2, we design a new encoder-decoder structure with an extra auxiliary learning part.

This module contains two parts, the attention generating part and the auxiliary learning part. The attention part uses a fully convolutional network [13] to generate the lesion area attention maps, containing a score value for each pixel, which indicates whether it refers to lesion part or normal skin part. So basically it is a binary segmentation task, which is supervised by normal cross-entropy loss, illustrated as L_{ce} in Fig. 2(b). After the attention map generated in the first part, the smallest rectangle which includes the lesion area is taken from the attention map to crop out the image. The cropped image is the output of this module, which is also the input of the next Feature Extraction Module. Meanwhile, the auxiliary learning part serves as an extra supervision for this module. The original image will be cropped by both the predicted attention map and the groundtruth map. The two cropped images are the inputs of this part and will then be fed to two convolutional blocks with same architecture, which is illustrated in Fig. 2(b). For

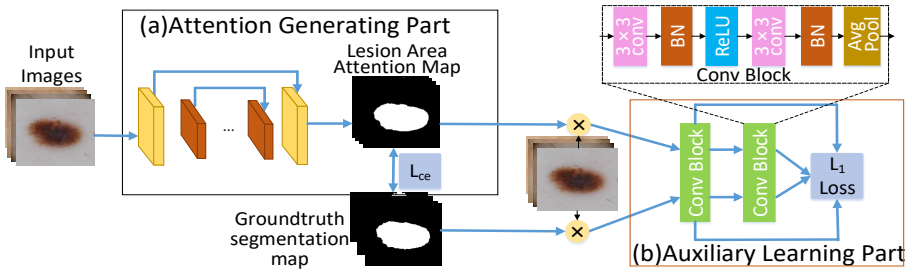

Fig. 2. The details of lesion area attention module.

each conv block, L_1 loss is used to compute the difference between the output features of the two branches. This loss is served as an extra regularization for the bourdaries since the mistakes may appear in these areas. To note that the auxiliary learning part will only be used at the training phase.

Feature Extraction Module This module is designed to simulate the **observe step** of doctors. Convolutional layers, as well as the activation layers are used in normal CNN to form this module. It takes the cropped images as the inputs and extracts the feature.

Lesion Feature Attention Module This module is proposed to simulate the **compare step** of doctors. As illustrated in Fig. 3, we design two blocks, the spatial attention block, as to enable the network to focus on important lesion regions, and the correlation attention block, as to mine the correlation between different lesion areas. Denote the input feature as F, the shape of which is $H \times W \times C$.

For the spatial attention block, F will first go through one 1×1 conv layer and two 3×3 conv layers along with a ReLU layer. The feature now is denoted as $\hat{F}$. Then, instead of the normal softmax layer, a spatial normalization function is conducted by Eqn. (1) to generate the attention map Q. Finally, an element-wise multiplication between Q and F is applied, then the result will be multiplied by a learnable weighting factor α and performed a summation to F.

$$Q = \{q|q_{ij} = \frac{\hat{F}_{ij}}{\sum_{i'=1}^{H}\sum_{j'=1}^{W}\hat{F}_{i'j'}}\} \tag{1}$$

For the relation attention block, the feature will first go through two branches. Each branch has one 1×1 conv layer and one 3×3 conv layer. The feature now is denoted as $\overline{F}$. Then $\overline{F}$ is reshaped to $M \times C$, where M is equal to $H \times W$. After that, a matrix multiplication between $\overline{F}$ and $\overline{F}^{'}$, the transpose of $\overline{F}$, is applied. The result is denoted as $\widetilde{F}$, which is a matrix of size $M \times M$. After that, the attention map P is generated by applying a normalization function on $\widetilde{F}$. P has a size of $M \times M$. The equation is:

$$\widetilde{F} = \overline{F} * \overline{F}^{'}, \qquad P = \{p|p_{ij} = \frac{exp(\widetilde{F}_{ij})}{\sum_{i'=1}^{M}\sum_{j'=1}^{M}exp(\widetilde{F}_{i'j'})}\} \tag{2}$$

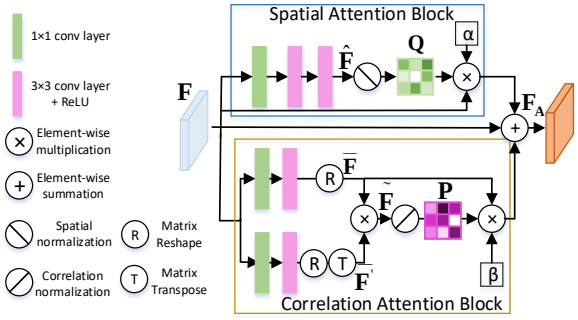

Fig. 3. Lesion feature attention module.

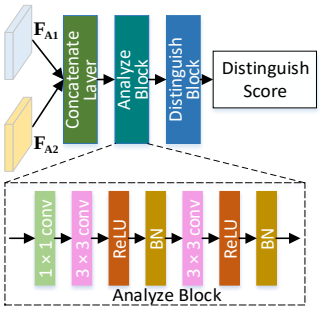

Fig. 4. Distinguish module.

where "*" means matrix multiplication, P is the attention map. p_{ij} represents the correlation between i^{th} position and j^{th} position in feature $\widetilde{F}$. After generating P, a matrix multiplication was applied with $\overline{F}$ and reshape the result to $H \times W \times C$. Then, the reshaped result will be multiplied by a learnable weighting factor β and applied a matrix summation to F. The final output of this module is obtained by Eqn. (3):

$$F_A = \alpha \cdot (Q \cdot F) + \beta \cdot (P * \overline{F}) + F \tag{3}$$

where "$\cdot$" means element-wise multiplication, "*" means matrix multiplication, α and β are scalars which are learnable weights, F_A is the output of the lesion feature attention module.

At the same time, F_A will be fed into two fully connected layers for the classification task, and supervised by weighted cross-entropy loss:

$$L_c = -\frac{1}{N_2} \sum_{i=1}^{N_2} \sum_{j=1}^{K} w_j (y_{ij} log(x_{ij})) \tag{4}$$

where N_2 is the number of images, K is the number of classes. w_j is the weight of class j. y_{ij} denotes whether image i belongs to label j, and x_{ij} means the classification score of class j for image i.

Distinguish Module This module is designed to mimic the **distinguish strategy** of doctors' learning process. Normally, there're three steps to distinguish whether the two images are from same class: integrate the important features of the them, analyze them, and draw conclusions. Inspired by this, we adopt three parts in this module: a concatenate layer, an analyze block and a distinguish block, illustrated in Fig. 4.

For the two input images, the feature of two images are first fed into the concatenated layer, in which a new concatenated feature is formed. Then the analyze block, which contains a 1×1 conv layer and two 3×3 conv layers, will mine a deeper feature reflecting the relationship between two images, which is illustrated by Fig. 4(a). Finally, a distinguish score indicating whether these two images are from the same class will be obtained by the distinguish block, which contains two fully connected layer with a ReLU and dropout layer between them.

A binary cross-entropy loss is used to supervise the module, illustrated as Eqn. (5):

$$L_d = -\frac{1}{N_1}\sum_{i=1}^{N_1}(y_d^i log(x_d^i) + (1-y_d^i)log(1-x_d^i)) \tag{5}$$

where N_1 denotes the number of image pairs, y_d^i is the distinguish label of the i_{th} image pair, and x_d^i is the distinguish score, predicted by the module.

This module is different from [22] in both perspective and architecture. The designing of the module is fully driven by the distinguish strategy in clinical process of doctors, including the three sub-steps in the strategy. However, the method in [22] is designed on the perpective of data, which is the intra-class variation and inter-class similarity in skin lesion data. Besides, the architecture is also different. The authors in [22] only adopt three single linear layers in the synergic network, in contrast we design two blocks, including several convolutional layers, activation layers and fully-connected layers, for a better non-linear learning ability of the distinguishment.

So the final loss of the whole network is:

$$L = L_c + \lambda L_d \tag{6}$$

where λ is a hyper-parameter to linearly combine two losses in order to balance them.

3 Experiments

3.1 Datasets

ISBI 2016 dataset [7] contains 900 images as training data and 379 images as testing data. The task is to classify images as either malignant (melanoma) or non-malignant (non-melanoma). ISIC 2017 dataset [3] contains 2000 training, 150 validation, and 600 test images. The dataset contains three classes, including melanoma, nevus and seborrheic keratosis. There are two binary classification sub-tasks, the first one is to distinguish between melanoma and the others (nevus and seborrheic keratosis), the second task is to distinguish between seborrheic keratosis and the others (nevus and melanoma). Since the groundtruth of the testset of ISIC2018 and 2019 dataset isn't available, we can't implement our method on these two datasets.

3.2 Implementation details

ResNet18 are used as the backbone network for feature extraction module. At the training phase, the lesion area attention module are trained seperately first. Since in ISBI2016 Challenge Part3B, all the other methods use the data with segmentation map to aid the classification task, we use these data to supervise the lesion area attention module. After training the lesion area attention module, two original images are firstly fed into the lesion area attention module and the cropped images are the input of later architecture. The other modules are trained by Eqn. (6) together. At the testing phase, a single image is fed into the lesion area attention module, the cropped image is used as the input of both branches. The probabilistic predictions obtained by each branch are

summed, and the class with highest probability will be the final result. The images are resized to 448 $\times$ 448 for training and testing. The learning rate is initialized to 0.001 and decayed by 0.1 every 20 epochs. The batch size is set to 32. The image pairs are randomly chosen. In Eqn. (6), λ is set to 1. Five evaluation metrics are used in the experiments, including Average Precision(AP), Accuracy(ACC), Area under the receiver operating characteristic curve(AUC), Sensitivity(SE) and Specificity(SP).

3.3 Ablation study

We conduct the ablation study on the second sub-task of ISBI2017 dataset to investigate the impact of the proposed three modules of our method on the performance. Note that during the challenge, it is ranked only by comparing the AUC value. The specific performance is listed in Table 3. The three main modules are listed in the following order: 1. Lesion Area Attention Module(**A**); 2. Lesion Feature Attention Module(**F**); 3. Distinguish Module(**D**). The feature extraction module is used by default. The baseline model is trained from pre-trained ResNet18, denoted as **"Baseline"**. We first compare the three modules when they are combined with the baseline respectively. The second, third, fourth rows are respectively the results of adding the lesion area attention module, as **"CIN(A)"**, the lesion feature attention module, as **"CIN(F)"** and the distinguish module, as **"CIN(D)"** to the baseline. The AUC value is improved by 1.7%, 1.3% and 1.2%, respectively. The experimental result proves that individually adopting the three modules can benefit the model, which reveals that our migration from the learning and diagnosis process to CNN is successful.

After that, we conduct the experiments with multiple modules added to the baseline. The performance of **CIN(A+F)** is better than **CIN(A)** and **CIN(F)**. The AUC value is further increased to 0.960, with a gain of 1.6% comparing to **CIN(A)**, and 2.0%, comparing to **CIN(F)**. Finally, when the Distinguish Module is introduced, it yields AUC of 0.981, which is 5.4% higher than that from the Baseline. The encouraging results show the advantage of our method. Supplementary material provides more ablation study results as well as the visualization results of Lesion Feature Attention Module.

Table 1. Ablation Study on ISIC2017 dataset sub-task2.

Method	AUC	ACC	SE	SP
Baseline	0.927	0.863	0.678	0.869
CIN(A)	0.944	0.922	0.778	0.947
CIN(F)	0.940	0.864	0.823	0.871
CIN(D)	0.939	0.905	0.711	0.939
CIN(A+F)	0.960	0.915	0.644	0.959
CIN (A+F+D)	**0.981**	**0.943**	**0.829**	**0.965**

Table 2. Performance of CIN and other methods on ISBI 2016 dataset.

Method	AP	ACC	AUC
Our method	**0.782**	**0.889**	**0.896**
CIN	0.740	0.887	0.873
DCNN-FV [21]	0.685	0.868	0.852
SDL [22]	0.664	0.858	0.818
CUMED [20]	0.637	0.855	0.804
GTDL [7]	0.619	0.813	0.802
Result2-3B [7]	0.615	0.844	0.808
USYD [7]	0.580	0.686	0.793
Mufic-IT [7]	0.534	0.760	0.685

Table 3. Ablation Study on ISIC2017 dataset sub-task2.

Method	AUC	ACC	SE	SP
Baseline	0.927	0.863	0.678	0.869
CIN(A)	0.944	0.922	0.778	0.947
CIN(F)	0.940	0.864	0.823	0.871
CIN(D)	0.939	0.905	0.711	0.939
CIN(A+F)	0.960	0.915	0.644	0.959
CIN (A+F+D)	**0.981**	**0.943**	**0.829**	**0.965**

Table 4. Performance of CIN and other methods on ISBI 2016 dataset.

Method	AC	SE	SP	AUC
Our method	**0.94**	0.925	**0.944**	**0.983**
MB-DCNN	**0.94**	0.95	0.938	0.977
CICS	-	**1**	0.882	-
MFLF	-	0.98	0.90	-
CCS	-	0.925	0.763	0.843

3.4 Results

We compare our performance to DCNN-FV [21], SDL [22] and the top5 methods of the challenge in 2016. To follow the tradition in the existing methods, we evaluate the performance on AP, ACC and AUC. Note that the challenge is ranked based only on **AP**, which means that it's the most important metric. The results are shown in Table 4 and our method are shown in the second row as **"CIN(ours)"**. We achieve the best result in all the three metric, with an AP of **0.740**, which exceeds DCNN-FV [21] by 5.5% and is almost 2.6 times of the difference between the next two methods. Our ACC and AUC outperforms the second place by 1.9% and 2.1%,respectively.

Note that both DCNN-FV [21] and SDL [22] use ResNet50 as the backbone but we achieves better results using ResNet18 with a smaller number of parameters.

Next, we compare our performance to [17], [23] and the top5 methods ISIC2017 challenge dataset. There are two sub-tasks in the challenge, to follow the tradition, we compared five metrics for each sub-task: AUC, AP, ACC, SE, and SP. Note that the challenge is ranked based only on the **Average AUC** of the two sub-tasks, which means that it's the most important metric. Table 5 shows the results. Since we use no extra images for the experiments, and couldn't find the same external data for [17, 23], we compare with the result of [17] without any extra image for a fair comparasion. It can be found from the table that the AUC achieved by our method of the two sub-tasks are both the highest, **0.920** and **0.981** respectively. The AUC of first sub-task improves by 3.7% and the second improves by 2.0%. The average AUC of our approach is **0.951**, which improves by 2.5% and is almost 3 times of the difference between the next two methods. Besides, the performance of AP and ACC in two sub-tasks is also the best. To sum up, our method not only achieves state-of-the-art performance among the rank metric of two challenge datasets with a significant improvement, but also gets the best results in most of other metrics, which fully demonstrates the advantages of our model. Supplymentary material provides more experiment results and visualization results.

4 Conclusion and Future Work

Motivated by the observations of learning and diagnosis of dermoscopy images from doctors, for the first time, we present a clinical-inspired network, to simulate the ac-

Table 5. Performance of CIN and other methods on ISIC 2017 dataset.

Methods	External data	Melanoma Classification					Seborrheic Keratosis					Average
		AUC	AP	ACC	SE	SP	AUC	AP	ACC	SE	SP	AUC
Our method	0	**0.947**	**0.868**	**0.903**	**0.803**	0.929	**0.986**	**0.929**	**0.947**	0.922	0.952	**0.967**
CIN	0	0.920	0.814	0.894	0.645	0.948	0.979	0.902	0.943	0.829	0.965	0.949
MB-DCNN	1320	0.903	-	0.878	0.727	0.915	0.973	-	0.93	0.844	0.945	0.938
ARL-CNN [23]	1320	0.875	-	0.850	0.658	0.896	0.958	-	0.868	0.878	0.867	0.917
SDL [9]	1320	0.868	0.689	0.872	-	-	0.955	0.818	0.917	-	-	0.912
RENI [11]	1444	0.868	0.710	0.828	0.735	0.851	0.953	0.786	0.803	**0.978**	0.773	0.911
gpm-LSSSD [8]	900	0.856	0.747	0.823	0.103	**0.998**	0.963	0.839	0.875	0.178	**0.998**	0.910
Alea-Jacta-Est [12]	7544	0.874	0.715	0.872	0.547	0.950	0.943	0.790	0.895	0.356	0.990	0.908
EResNet [1]	1600	0.870	0.732	0.858	0.427	0.963	0.921	0.770	0.918	0.589	0.976	0.896
DLSL [4]	1341	0.836	0.665	0.845	0.350	0.965	0.935	0.808	0.913	0.556	0.976	0.886

tual subjective process. The network includes a lesion area attention module and lesion feature attention module as to mimic the diagnosis process of doctors. Besides, the introduction of distinguish module are based on the original intention of simulating the learning process from doctors. At the same time, our method achieves encouraging performance while keeping a simple backbone network, which contains smaller parameters, and utilize no external data. The state-of-the-art experiment results on ISBI 2016 and ISIC 2017 datasets verify the effectiveness and advancement of our methods. In future works, we will try to integrate the lesion area attention module in the whole workflow.

Acknowledgement. This work was partially supported by the Natural Science Foundation of China under contracts 61572042 and 61772041. We also acknowledge the Clinical Medicine Plus X-Young Scholars Project, and High-Performance Computing Platform of Peking University for providing computational resources.

References

1. Bi, L., Kim, J., Ahn, E., Feng, D.: Automated skin lesion analysis using large-scale dermoscopy images and deep residual networks. arXiv preprint arXiv:1703.04197 (2017)
2. Binder, M., Schwarz, M., Winkler, A., Steiner, A., Kaider, A., Wolff, K., Pehamberger, H.: Epiluminescence microscopy: a useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists. Archives of dermatology **131**(3), 286–291 (1995)
3. Codella, N.C., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., 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). pp. 168–172. IEEE (2018)
4. DeVries, T., Ramachandram, D.: Skin lesion classification using deep multi-scale convolutional neural networks. arXiv preprint arXiv:1703.01402 (2017)
5. Ge, Z., Demyanov, S., Bozorgtabar, B., Abedini, M., Chakravorty, R., Bowling, A., Garnavi, R.: Exploiting local and generic features for accurate skin lesions classification using clinical and dermoscopy imaging. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). pp. 986–990. IEEE (2017)

6. Ge, Z., Demyanov, S., Chakravorty, R., Bowling, A., Garnavi, R.: Skin disease recognition using deep saliency features and multimodal learning of dermoscopy and clinical images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 250–258. Springer (2017)
7. Gutman, D., Codella, N.C., Celebi, E., Helba, B., Marchetti, M., Mishra, N., Halpern, A.: Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic). arXiv preprint arXiv:1605.01397 (2016)
8. Iv: Incorporating the knowledge of dermatologists to convolutional neural networks for the diagnosis of skin lesions. arXiv preprint arXiv:1703.09176 (2017)
9. Jianpeng, Z., Yutong, X., Qi, W., Yong, X.: Medical image classification using synergic deep learning. Medical Image Analysis **54**
10. Kittler, H., Pehamberger, H., Wolff, K., Binder, M.: Diagnostic accuracy of dermoscopy. The lancet oncology **3**(3), 159–165 (2002)
11. Matsunaga, K., Hamada, A., Minagawa, A., Koga, H.: Image classification of melanoma, nevus and seborrheic keratosis by deep neural network ensemble. arXiv preprint arXiv:1703.03108 (2017)
12. Menegola, A., Tavares, J., Fornaciali, M., Li, L.T., Avila, S., Valle, E.: RECOD titans at isic challenge 2017. arXiv preprint arXiv:1703.04819 (2017)
13. Olaf, R., Philipp, F., Thomas, B.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241 (2015)
14. Rogers, H.W., Weinstock, M.A., Feldman, S.R., Coldiron, B.M.: Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the US population, 2012. JAMA dermatology **151**(10), 1081–1086 (2015)
15. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics, 2015. CA: a cancer journal for clinicians **65**(1), 5–29 (2015)
16. Silveira, M., Nascimento, J.C., Marques, J.S., Marçal, A.R., Mendonça, T., Yamauchi, S., Maeda, J., Rozeira, J.: Comparison of segmentation methods for melanoma diagnosis in dermoscopy images. IEEE Journal of Selected Topics in Signal Processing **3**(1), 35–45 (2009)
17. Wenbo, Z., Chao, G., Lan, Y.: A relation hashing network embedded with prior features for skin lesion classification. In: International Conference on Medical image computing and computer-assisted intervention. pp. 115–123 (2018)
18. Yang, J., Sun, X., Liang, J., Rosin, P.L.: Clinical skin lesion diagnosis using representations inspired by dermatologist criteria. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1258–1266 (2018)
19. Yoon, C., Hamarneh, G., Garbi, R.: Generalizable feature learning in the presence of data bias and domain class imbalance with application to skin lesion classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 365–373. Springer (2019)
20. Yu, L., Chen, H., Dou, Q., Qin, J., Heng, P.A.: Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Transactions on Medical Imaging **36**(4), 994–1004 (2017)
21. Yu, Z., Jiang, X., Zhou, F., Qin, J., Ni, D., Chen, S., Lei, B., Wang, T.: Melanoma recognition in dermoscopy images via aggregated deep convolutional features. IEEE Transactions on Biomedical Engineering **66**(4), 1006–1016 (2019)
22. Zhang, J., Xie, Y., Wu, Q., Xia, Y.: Skin lesion classification in dermoscopy images using synergic deep learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 12–20. Springer (2018)
23. Zhang, J., Xie, Y., Xia, Y., Shen, C.: Attention residual learning for skin lesion classification. IEEE Transactions on Medical Imaging **38**(9), 2092–2103 (2019)