Soft Attention Improves Skin Cancer Classification Performance

Soumyya Kanti Datta[0000−0002−4219−5163], Mohammad Abuzar Shaikh[0000−0002−4235−6246], Sargur N. Srihari, and Mingchen Gao[0000−0002−5488−8514]

State University of New York at Buffalo, USA
{soumyyak,mshaikh2,srihari,mgao8}@buffalo.edu

Abstract. In clinical applications, neural networks must focus on and highlight the most important parts of an input image. Soft-Attention mechanism enables a neural network to achieve this goal. This paper investigates the effectiveness of Soft-Attention in deep neural architectures. The central aim of Soft-Attention is to boost the value of important features and suppress the noise-inducing features. We compare the performance of VGG, ResNet, Inception ResNet v2 and DenseNet architectures with and without the Soft-Attention mechanism, while classifying skin lesions. The original network when coupled with Soft-Attention outperforms the baseline[16] by 4.7% while achieving a precision of 93.7% on HAM10000 dataset [25]. Additionally, Soft-Attention coupling improves the sensitivity score by 3.8% compared to baseline[31] and achieves 91.6% on ISIC-2017 dataset [2]. The code is publicly available at github[1].

1 Introduction

Skin cancer is the most common cancer and one of the leading causes of death worldwide. Every day, more than 9500 people[2] in the United States are diagnosed with skin cancer, with 3.6 million people[3] diagnosed with basal cell skin cancer each year. Early diagnosis of the illness has a significant effect on the patients' survival rates. As a result, detecting and classifying skin cancer is important.

It is difficult to distinguish between malignant and benign skin diseases because they look so similar. Although a dermatologist's visual examination is the first step in detecting and diagnosing a suspicious skin lesion, it is usually followed by dermoscopy imaging for further analysis [32]. Dermoscopy images provide a high-resolution magnified image of the infected skin region, but they are not without their drawbacks. Due to the image size being large, it becomes difficult for the feature extractors to extract out the relevant features for classification. Various methods such as Segmentation and detection, Transfer learning, General Adversarial networks, etc. have been used to detect and classify skin cancer. Despite significant progress, skin cancer classification is still a difficult task. This is due to the lack of annotated data and low inter-class variation.

[1] https://github.com/skrantidatta/Attention-based-Skin-Cancer-Classification
[2] https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/
[3] https://www.skincancer.org/skin-cancer-information/basal-cell-carcinoma/

Furthermore, the task is complicated by contrast variations, color, shape, and size of the skin lesion, as well as the presence of various artifacts such as hair and veins. Inspired by the work done in [18], this paper studies the effect of soft attention mechanism in deep neural networks. Deep learning architectures identify the image class by learning the salient features and nonlinear interactions. The soft-attention mechanism improves performance by focusing primarily on relevant areas of the input. Moreover, the soft-attention mechanism makes the image classification process transparent to medical personnel, as it maps the parts of the input that the network uses to classify the image, thereby, increasing trust in the classification model.

Following Krichevsky[12], large-scale image classification tasks using deep convolutional neural networks have become common. As reported in the paper[3], the task of skin cancer classification using images has improved rapidly since the implementation of Deep Neural Networks. To make progress, we suggest that soft attention be used to identify fine-grained variability in the visual features of skin lesions.

Existing art in the field of skin cancer classification used streamlined pipelines based upon current Computer Vision [4]. Masood et al. in their paper [13] proposed a general framework from the viewpoint of computer vision, where the methods such as calibration, preprocessing, segmentation, balancing of classes and cross validation are used for automated melanoma screening. In 2018, Valle et al. [26] investigated ten different methodologies to evaluate deep learning models for skin lesion classification. Data augmentation, model architecture, image resolution, input normalization, train dataset, use of segmentation, test data augmentation, additional use of support vector machines, and use of transfer learning are among the ten methodologies they evaluated. They stated that data augmentation had the greatest impact on model efficiency. The same observation is confirmed by Perez's 2018 paper "Data Augmentation for Skin Lesion Analysis"[15].

Nonetheless, the problems of low inter-class variance and class imbalance in skin lesion image datasets remain, seriously limiting the capabilities of deep learning models[30]. To fix the lack of annotated data, Zunair et al.[32] proposed the use of adversarial training and Bissoto et al.[1] proposed the use of Generative Adversarial Networks to produce realistic synthetic skin lesion photos. Zhang et al. [31] in 2019 proposed the attention residual learning convolutional neural network for skin lesion classification which is based on self attention mechanism.

In this paper, we suggest using a Soft Attention mechanism in conjunction with a Deep Convolutional Neural Network (DCNN) to classify skin cancer. Rather than using attention modules with residual blocks and stacking them one after another like in paper [31], we integrated the soft attention module into the various DCNN architectures such as Inception ResNet v2[22], which improved the performance of those architectures. Our model used the DCNN to extract the features maps from the skin lesion images, and the soft attention module assisted the DCNN in focusing more on the important features of the images without completely discarding the other features. Our paper's primary contribution is that we offer a unique technique for integrating soft attention mechanism with DCNNs to optimize performance, and we outperformed the state-of-the-art on skin lesion classification on the HAM10000 dataset[25] and ISIC-2017 dataset [2].

2 Method

In this paper, five deep neural networks which are ResNet34, ResNet50 [6], Inception ResNet v2[22], DenseNet201[8] and VGG16 [20], are implemented with soft attention mechanism, to classify skin cancer images. ResNet34, ResNet50[6], Inception ResNet v2, DenseNet201[8] and VGG16[20] are all state of the art feature extractors which are trained on ImageNet dataset. The main components and architecture of the proposed approach is described below:

2.1 Dataset

The experiment is performed on two datasets separately. The two datasets are as follows: HAM10000 dataset [25] and ISIC 2017 dataset [2].

The HAM10000 dataset [25] consists of 10015 dermatoscopic images of a size of 450×600. It consists of 7 diagnostic categories as follows: Melanoma(MEL), Melanocytic Nevi(NV), Basal Cell Carcinoma(BCC), Actinic Keratosis, and Intra-Epithelial Carcinoma(AKIEC), Benign Keratosis(BKL), Dermatofibroma(DF), Vascular lesions(VASC). All the images are resized to 299×299 for Inception ResNet v2[22] architecture and 224×224 for the other architectures.

The ISIC 2017 dataset consists of 2600 images. In the training dataset there are 2000 images of 3 catagories as follows: benign nevi, seborrheic keratosis, and melanoma. The test dataset consist of 600 images. In this experiment we are training our model to classify only benign nevi and seborrheic keratosis. All the images resized to 224×224.

The data in both datasets is then cleaned to remove class imbalances. This is done by the process of over-sampling and under-sampling of data so that there are equal number of images per class. The images are then normalized by dividing each pixel with 255 to keep the pixel values in the range 0 to 1.

2.2 Soft Attention

When it comes to skin lesion images, only a small percentage of pixels are relevant as the rest of the image is filled with various irrelevant artifacts such as veins and hair. So, to focus more on these relevant features of the image, soft attention is implemented. Inspired by the work proposed by Xu et al [28], for image caption generation and the work done by Shaikh et al [18], where they used attention mechanism on images for handwriting verification, in this paper, soft attention is used to classify skin cancer.

In Figure [2], we can see that areas with higher attention are red in color. This is because soft attention discredits irrelevant areas of the image by multiplying the corresponding feature maps with low weights. Thus the low attention areas have weights closer to 0. With more focused information, the model performs better.

In the soft attention module as discussed in paper [18] and [23], the feature tensor (t) which flows down the deep neural network is used as input.

$$f_{sa} = \gamma t((\sum_{k=1}^{K} softmax(W_k * t))) \tag{1}$$

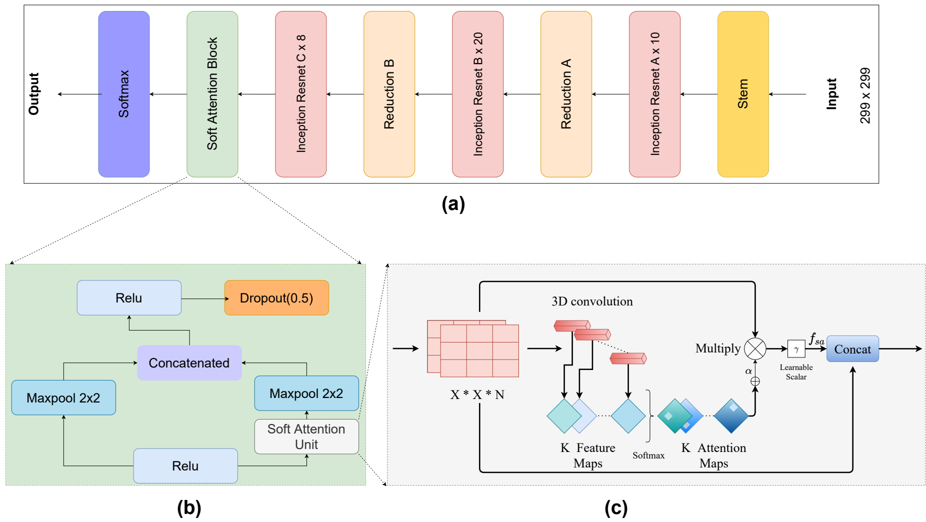

Fig. 1. (a). End to end architecture Of Inception ResNet v2[22] with Soft Attention Block. **(b).** The schema for Soft Attention Block. **(c).** Soft Attention unit

This feature tensor $t \in \mathbb{R}^{h \times w \times d}$ is input to a 3D convolution layer[24] with weights $W_k \in \mathbb{R}^{h \times w \times d \times K}$, where K is the number of 3D weights. The output of this convolution is normalized using softmax function to generate $K = 16$ attention maps. As shown in Figure 1(c), these attention maps are aggregated to produce a unified attention map that acts as a weighting function α. This α is then multiplied with t to attentively scale the salient feature values, which is further scaled by γ a learnable scalar. Because various images require different γ values, γ is treated as a learnable parameter and not hard-coded. This allows the network to determine how much it should focus on the attention maps on its own. Finally, the attentively scaled features (f_{sa}) are concatenated with the original feature t in form of a residual branch. During training we initialize γ from 0.01 so that the network can slowly learn to regulate the amount of attention required by the network.

2.3 Model Setup

In this section, the detailed architecture of the models is discussed. For all experiments, to train the networks, Adam optimizer[11] of 0.01 learning rate and 0.1 epsilon is used. A batch normalization[10] layer is added after each layer in all the networks to introduce some regularization. For the HAM10000 dataset [25], since there are 7 classes of skin cancer, an output layer with 7 hidden units is implemented, followed by a softmax activation unit. We employed a batch size of 16 during both training and testing. In this paper, the model is evaluated using $Precision = \frac{TP}{TP+FP}$, $Accuracy = \frac{TP+TN}{T}$, $Sensitivity = \frac{TP}{TP+FN}$, $Specificity = \frac{TN}{TN+FP}$ and AUC scores[9]. Here TN, TP, FP, FN, T mean, True Negatives, True Positives, False Positives, False Negatives, Total

Number respectively. All the experiments were executed on the Keras framework with tensorflow version 2.4.0.

Inception ResNet v2: In Inception ResNet v2[22], the soft attention layer is added to the Inception Resnet C block of the model where the feature size of the image is 8×8 as shown in Figure [1(a)]. In this case, the soft attention layer is followed by a maxpool layer with a pool size of 2×2, which is then concatenated with the filter concatenate layer of the inception block. The concatenate layer is then followed by a relu activation unit. To regularize the output of the attention layer, the activation unit is followed by a dropout layer[21] with dropout probability of 0.5 as shown in Figure [1(b)]. The network is trained for 150 epochs with early stopping patience of 30 epochs while monitoring the validation loss for a minimum delta of 0.001. The dropout and early stopping regularization prevents the network from over-fitting to the training dataset. The overall network is shown in Figure [1(a)]. The other architectures, such as ResNet34, ResNet50 [6], DenseNet201[8] , and VGG16[20], and the process by which the Soft Attention block was integrated with them, are described in the supplementary document.

2.4 Loss Function

In this experiment, there are seven different classes of skin cancer. Hence , categorical cross entropy loss (L_{CCE}) is used to optimize the neural network.

$$L_{CCE} = -\sum_{i=1}^{C} t_i log(f(s)_i) \tag{2}$$

where

$$f(s)_i = \frac{e^{s_i}}{\sum_{j=1}^{C} e^{s_j}} \tag{3}$$

Here, as there are seven classes, $C \in [0..6]$, where t_i is the ground truth and s_i is the CNN score for each class i in C. $f(s)_i$ is the softmax activation function applied to the scores.

3 Results

3.1 Ablation Analysis

Table 1 lists, the performance of all the models in terms of precision, and AUC score on HAM10000 dataset [25]. In this table (+SA) stands for models with soft attention. IRv2 stands for Inception ResNet v2[22], [6]34 stands for ResNet34[6] and [6]50 stands for ResNet50[6]. From the table, it can be observed that IRv2 when coupled with SA (IRv2[22]+SA) shows significant improvements in results, with a precision and AUC score of 93.7% and 98.4% respectively, which are also the highest scores amongst all models. Furthermore, we can see that Soft Attention (SA) boosts the performance of

Dis.	Precision										AUC										#
	[22]	[22] + SA	[8]	[8] + SA	[20]	[20] + SA	[6]50	[6]50 + SA	[6]34	[6]34 + SA	[22]	[22] + SA	[8]	[8] + SA	[20]	[20] + SA	[6]50	[6]50 + SA	[6]34	[6]34 + SA	
AKIEC	0.830	**1.000**	**1.000**	0.920	0.620	0.700	0.740	0.670	0.670	0.500	**0.993**	0.981	0.975	0.967	0.949	0.964	0.980	0.981	0.969	0.970	23
BCC	0.850	0.880	0.830	0.800	0.540	0.620	**0.910**	0.880	0.660	0.880	0.997	**0.998**	0.993	0.994	0.977	0.984	0.997	0.996	0.991	0.993	26
BKL	**0.850**	0.720	0.690	0.730	0.570	0.630	0.670	0.670	0.510	0.520	0.970	**0.982**	0.960	0.964	0.930	0.900	0.948	0.964	0.904	0.916	66
DF	0.670	**1.000**	0.500	**1.000**	0.250	0.500	0.800	**1.000**	0.400	0.330	0.973	**0.982**	0.851	0.921	0.847	0.809	0.973	0.971	0.925	0.949	6
MEL	0.700	0.670	0.540	0.530	0.500	0.430	0.520	**0.730**	0.420	0.540	0.965	0.974	0.963	**0.976**	0.925	0.956	0.961	0.973	0.910	0.953	34
NV	0.930	**0.970**	0.950	0.950	0.930	0.950	0.950	0.950	0.930	0.930	**0.984**	**0.984**	0.975	0.976	0.954	0.951	0.974	0.979	0.944	0.958	663
VASC	**1.000**	**1.000**	0.900	0.830	**1.000**	**1.000**	0.900	**1.000**	0.910	0.820	**1.000**	**1.000**	0.993	0.999	0.972	0.999	0.995	0.999	0.999	0.996	10
Avg	0.832	**0.892**	0.771	0.824	0.631	0.690	0.783	0.841	0.642	0.646	0.983	**0.984**	0.959	0.971	0.936	0.937	0.975	0.980	0.949	0.962	828
W. Avg	0.905	**0.937**	0.904	0.909	0.862	0.882	0.898	0.910	0.857	0.865	0.982	**0.984**	0.974	0.975	0.951	0.948	0.972	0.978	0.942	0.957	828

Table 1. Ablation results for choosing the best model on HAM10000 dataset [25]. [22] refers to IRv2 architecture, [8] refers to DenseNet 201 architecture, [20] refers to VGG 16 architecture, and [6] refers to ResNet architecture.

IRv2 by 3.2% in terms of precision as compared to the original IRv2 model. This phenomenon is true for VGG16, ResNet34, ResNet50 and DenseNet201 as well. For instance, Soft Attention (SA) boosts the precision of DenseNet201[8], ResNet34[6], ResNet50[6], and VGG16[20] by 0.5%, 0.8%, 1.2% and 2% respectively. We see a similar behaviour for the AUC scores when SA block is integrated in to the networks, such as, the performance of ResNet50[6], and ResNet34[6] has grown by 0.6% and 1.5% respectively and the performance of DenseNet201[8], and VGG16[20] is on par with the original models.

Although IRv2+SA performs the best in terms of weighted average(W.Avg), when we look at it's class wise performance, we can see that Soft Attention enhances the efficiency of the original IRv2 while categorizing AKIEC, BCC, DF and NV by 17%, 3%, 33% and 4% respectively in terms of precision. Moreover, when comparing AUC scores, the IRv2+SA performs better for BKL and MEL by 1.2% and 0.9% respectively, while, for BCC, NV and VASC, IRv2+SA performs as good as original model.

We thus select IRv2 coupled with SA (IRv2+SA) for our experiments, also the SA block consistently boosts the performance of it's original counterpart, hence, we can justify the integration of Soft Attention to the networks.

3.2 Quantitative Analysis

The proposed approach is compared with state-of-the-art models for skin cancer classification on the HAM10000 dataset [25] in Table 2(a). Our Soft Attention-based approach outperforms the baseline[16] by 4.7% in terms of precision. In terms of AUC scores, our Soft Attention-based approach clearly outperforms them all by 0.5% to 4.3%.

We also tested the model with different train-test splits on the HAM10000 dataset [25], we discovered that the model with 85% training data outperforms the model with 80% and 70% training data by 2.2% and 2.6% respectively, as shown in supplementary material's Table 1. Hence we select 85/15% training/testing split for performing our experiments. In Table 2(b), the performance of the proposed approach Inception Resnet V2[22] (IRv$2_{5\times5}$+SA and IRv$2_{12\times12}$+SA) with soft attention is measured on

Model	Avg AUC	Precision	Accuracy
Loss balancing and ensemble[5]	0.941	-	0.926
Single Model Deep Learning[29]	0.974	-	0.864
Data classification augmentation[19]	0.975	-	0.853
Two path CNN model[14]	-	-	0.886
Various Deep CNN (Baseline) [16]	0.979	0.890	-
IRv2+SA(Proposed Approach)	**0.984**	**0.937**	**0.934**

(a)

Networks	AUC	Accuracy	Sensitivity	Specificity
ResNet50 [6]	0.948	0.842	0.867	0.837
RAN50 [27]	0.942	0.862	0.878	0.859
SEnet50 [7]	0.952	0.863	0.856	0.865
ARL-CNN50 [31]	0.958	0.868	0.878	**0.867**
($IRv2_{12x12}$+SA)vs. N&M	0.922	0.890	**0.956**	0.589
($IRv2_{5x5}$+SA)vs. N&M	0.942	0.900	0.932	0.687
($IRv2_{12x12}$+SA)vs. N	0.935	0.898	0.945	0.711
($IRv2_{5x5}$+SA)vs. N	**0.959**	**0.904**	0.916	0.833

(b)

Table 2. (a). Comparison with state-of-the-art-Model in terms of Average AUC score, Precision and Accuracy on HAM10000 dataset [25]. **(b).** Comparison with state-of-the-art-Model in terms of AUC, Accuracy, sensitivity and specificity score on ISIC-2017 dataset [2] for Seborrheic Keratosis classification

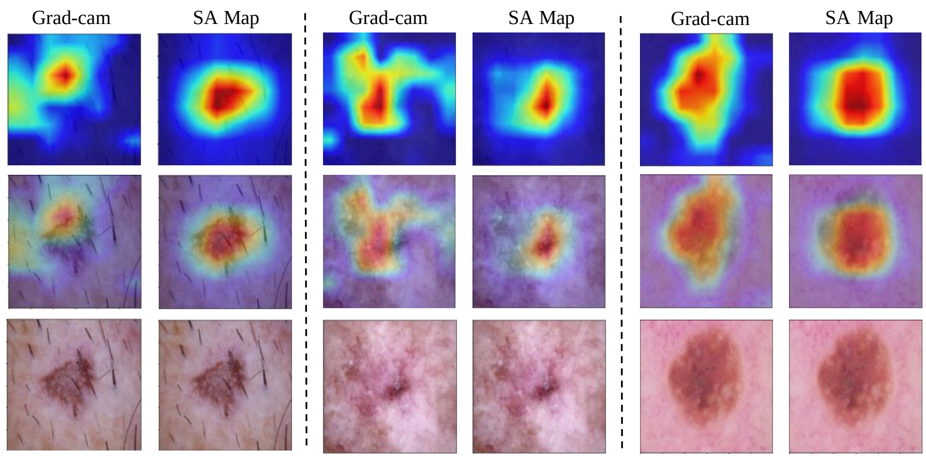

Fig. 2. Comparison of GradCAM [17] heatmaps with our Soft Attention (SA) maps on HAM10000 dataset [25]

ISIC-2017 dataset [2] on basis of AUC scores, Accuracy , Sensitivity and Specificity with the state-of-the-art models. Here ($IRv2_{5\times5}$+SA vs. N and $IRv2_{12\times12}$+SA vs. N) refers to classification of seborrheic keratosis with respect to only benign nevi, and ($IRv2_{5\times5}$+SA vs. N&M and $IRv2_{12\times12}$+SA vs. N&M) refers to classification of seborrheic keratosis with respect to both benign nevi and melanoma.

From Table 2(b), it can be observed that in $IRv2_{5\times5}$+SA, and in $IRv2_{12\times12}$+SA, the attention layer was added when the feature map size is 5×5 and 12×12 respectively. Out of the four models with soft attention, the model $IRv2_{5\times5}$+SA vs. N outperforms $IRv2_{12\times12}$+SA vs. N, $IRv2_{5\times5}$+SA vs. N&M and $IRv2_{12\times12}$+SA vs. N&M , in terms of AUC scores by 1.7% to 3.7% , Accuracy by 0.4% to 1.4%, and Specificity by a percentage of 12.2% to 24.4% respectively whereas $IRv2_{12\times12}$+SA vs. N&M outperforms $IRv2_{5\times5}$+SA vs. N, $IRv2_{5\times5}$+SA vs. N&M and $IRv2_{12\times12}$+SA vs. N in terms of Sensitivity by 4.0%, 2.4% and 1.1% respectively. When $IRv2_{5\times5}$+SA vs. N is compared with the ARL-CNN50[31] (baseline model), it performs on par with it in terms of AUC score but our model outperforms it when it comes to accuracy and Sensitivity by 3.6% and 3.8% respectively. But ARL-CNN50[31] takes the upper hand when it comes to Specificity by 3.4%. Since sensitivity measures the proportion of correctly identified positives and specificity measures the proportion of correctly identified negatives, we are prioritizing Sensitivity because classifying a person with cancer as not having cancer is riskier than vice versa.

3.3 Qualitative Analysis

Fig.2 displays the Soft Attention heat maps from the IRv2+SA model. In the Fig.2, the images on the bottom row are the input images from the HAM10000 dataset[25]. The images in the middle row under the columns SA Map show the Soft Attention maps superimposed on input images to show where the model is focusing and the images of the top row are attention maps themselves.

In Fig.2, we show pairs of comparison between the Soft Attention maps with Grad-CAM [17] heatmaps. In the first pair, the SA map focuses on the main part of the lesion area whereas the Grad-cam heatmap is slightly shifted towards top left and is also spread out on the uninfected area of skin. We have similar observations for the second and third pairs as well. From this observation it is evident that the Soft Attention maps are focused more on the relevant locations of the image compared to Grad-CAM[17] heatmaps.

4 Conclusion

In this paper, we present the implementation and utility of Soft Attention mechanism being applied while image encoding to tackle the problem of high-resolution skin cancer image classification. The model outperformed the current state-of-the-art approaches on the HAM10000 dataset [25] and the ISIC-2017 dataset [2]. This demonstrates the Soft Attention based deep learning architecture's potential and effectiveness in image analysis as well as in skin cancer classification. The Soft Attention mechanism also eliminates the need of using external mechanisms like GradCAM [17], and internally provides the location of where the model focuses while categorizing a disease, while also boosting the performance of the main network. Soft Attention has the added advantage of naturally dealing with image noise internally.

In future, we believe the salient regions proposed by Soft Attention can be used as salient regions for downstream tasks like classification, Visual Q&A and captioning as it will benefit datasets that don't have any bounding box annotations. Furthermore, this model can be also implemented in dermoscopy systems to assist dermatologists. Lastly, this mechanism can easily be implemented to classify data from other medical databases as well.

References

1. Bissoto, A., Perez, F., Valle, E., Avila, S.: Skin lesion synthesis with generative adversarial networks. In: OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis, pp. 294–302. Springer (2018)

2. Codella, N.C.F., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N.K., Kittler, H., Halpern, A.: 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). CoRR **abs/1710.05006** (2017), http://arxiv.org/abs/1710.05006
3. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S.: Dermatologist-level classification of skin cancer with deep neural networks. nature **542**(7639), 115–118 (2017)
4. Fornaciali, M., Carvalho, M., Bittencourt, F.V., Avila, S., Valle, E.: Towards automated melanoma screening: Proper computer vision & reliable results. arXiv preprint arXiv:1604.04024 (2016)
5. Gessert, N., Nielsen, M., Shaikh, M., Werner, R., Schlaefer, A.: Skin lesion classification using ensembles of multi-resolution efficientnets with meta data. MethodsX p. 100864 (2020)
6. 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. pp. 770–778 (2016)
7. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132–7141 (2018)
8. 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. pp. 4700–4708 (2017)
9. Huang, J., Ling, C.X.: Using auc and accuracy in evaluating learning algorithms. IEEE Transactions on knowledge and Data Engineering **17**(3), 299–310 (2005)
10. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)
11. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
12. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Communications of the ACM **60**(6), 84–90 (2017)
13. Masood, A., Ali Al-Jumaily, A.: Computer aided diagnostic support system for skin cancer: a review of techniques and algorithms. International journal of biomedical imaging **2013** (2013)
14. Nadipineni, H.: Method to classify skin lesions using dermoscopic images. arXiv preprint arXiv:2008.09418 (2020)
15. Perez, F., Vasconcelos, C., Avila, S., Valle, E.: Data augmentation for skin lesion analysis. In: OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis, pp. 303–311. Springer (2018)
16. Rezvantalab, A., Safigholi, H., Karimijeshni, S.: Dermatologist level dermoscopy skin cancer classification using different deep learning convolutional neural networks algorithms. arXiv preprint arXiv:1810.10348 (2018)
17. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. pp. 618–626 (2017)
18. Shaikh, M.A., Duan, T., Chauhan, M., Srihari, S.N.: Attention based writer independent verification. 2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR) (Sep 2020). https://doi.org/10.1109/icfhr2020.2020.00074, http://dx.doi.org/10.1109/ICFHR2020.2020.00074
19. Shen, S., Xu, M., Zhang, F., Shao, P., Liu, H., Xu, L., Zhang, C., Liu, P., Zhang, Z., Yao, P., et al.: Low-cost and high-performance data augmentation for deep-learning-based skin lesion classification. arXiv preprint arXiv:2101.02353 (2021)

20. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
21. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research **15**(1), 1929–1958 (2014)
22. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.: Inception-v4, inception-resnet and the impact of residual connections on learning. arXiv preprint arXiv:1602.07261 (2016)
23. Tomita, N., Abdollahi, B., Wei, J., Ren, B., Suriawinata, A., Hassanpour, S.: Attention-based deep neural networks for detection of cancerous and precancerous esophagus tissue on histopathological slides. JAMA network open **2**(11), e1914645–e1914645 (2019)
24. Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M.: Learning spatiotemporal features with 3d convolutional networks. In: Proceedings of the IEEE international conference on computer vision. pp. 4489–4497 (2015)
25. Tschandl, P., Rosendahl, C., Kittler, H.: The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific data **5**(1), 1–9 (2018)
26. Valle, E., Fornaciali, M., Menegola, A., Tavares, J., Bittencourt, F.V., Li, L.T., Avila, S.: Data, depth, and design: Learning reliable models for skin lesion analysis. Neurocomputing **383**, 303–313 (2020)
27. Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., Wang, X., Tang, X.: Residual attention network for image classification (2017)
28. Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., Zemel, R., Bengio, Y.: Show, attend and tell: Neural image caption generation with visual attention. In: International conference on machine learning. pp. 2048–2057 (2015)
29. Yao, P., Shen, S., Xu, M., Liu, P., Zhang, F., Xing, J., Shao, P., Kaffenberger, B., Xu, R.X.: Single model deep learning on imbalanced small datasets for skin lesion classification. arXiv preprint arXiv:2102.01284 (2021)
30. 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 (2016)
31. 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)
32. Zunair, H., Hamza, A.B.: Melanoma detection using adversarial training and deep transfer learning. Physics in Medicine & Biology (2020)