Depthwise Separable Convolutional Neural Network for Skin Lesion Classification

Sara Hosseinzadeh Kassani
Department of Computer Science
University of Saskatchewan
Saskatoon, Canada
sara.kassani@usask.ca

Peyman Hosseinzadeh Kassani
Department of Biomedical Engineering
University of Tulane
New Orleans, USA
peymanhk@tulane.edu

Michal J. Wesolowski
Department of Medical Imaging
University of Saskatchewan
Saskatoon, Canada
mike.wesolowski@usask.ca

Kevin A. Schneider
Department of Computer Science
University of Saskatchewan
Saskatoon, Canada
kevin.schneider@usask.ca

Ralph Deters
Department of Computer Science
University of Saskatchewan
Saskatoon, Canada
deters@cs.usask.ca

***Abstract*—Melanoma is one of the deadliest skin cancers. Early diagnosis plays an essential role in effective treatment planning and reducing the mortality rate of skin cancer. In this study, we propose a compact deep learning-based classification model with a separable convolutional neural network for melanoma detection. The proposed architecture is aimed to minimize the need for task-specific data pre-processing methods such as noise reduction, artifact removal, and low contrast adjustment to support a better generalization ability for unseen / test data. We validated the performance of the proposed method on the ISIC 2018 dataset. Our results show that the proposed architecture achieves comparable accuracy to the widely-used architectures presented in the literature while being more compact and simpler. The proposed methodology achieves 87.24% accuracy, 95.94% sensitivity and 98.47% specificity.**

***Index Terms*—Deep learning, Medical imaging, Melanoma detection, Separable convolution, Skin cancer classification**

I. Introduction

Melanoma develops from melanin producing melanocyte cells. It is one of the most aggressive types of skin cancer. According to the annual report provided by American cancer society [1], approximately 96,480 new cases will be diagnosed (about 57,220 in men and 39,260 in women) in the United States, and 7230 cases are expected to die in 2019. Early detection of melanoma can help improve patient survival rates. However, accurate and early detection of skin cancer can be a challenging task due to variations in shape, size, noisy artifacts and heterogeneous textures. With the rapid development of computer-aided diagnosis (CAD) systems in recent years, deep learning with convolutional neural network (CNN) has been widely used for different tasks such as classification [2]–[4], segmentation [5], and object detection [6].

Dermoscopy is a non-invasive and inexpensive tool that is used to capture high-quality images to support healthcare professionals in skin lesions examination. Employing CAD systems with the ability to analyze dermoscopic images may complement the clinical assessment of dermatologists for early and accurate melanoma detection [7]. In the next section, we briefly review the recent works on skin cancer classification and present the motivation for developing a depthwise separable convolutional neural network for skin cancer classification.

A. Related studies

Melanoma detection has been widely investigated [8]–[11]. For brevity, only research based on the ISIC 2018 dataset will be discussed here-in. An ensemble of Squeeze-and-Excitation Networks (SENet) and semi-supervised learning methods obtained promising results in a system proposed by Kitada et al. [12]. The authors developed specially designed data augmentation strategies such as between-class learning [13], random erasing [14] and body hair augmentation to improve the accuracy in addition to the commonly used data augmentation methods such as flipping, shearing, rotating. Kitada's method achieved 87.20% accuracy on the ISIC 2018 dataset. In a study by Gessert et al. [15], an ensemble model based on fine-tuning of pre-trained DenseNet [16], SENet [17] and ResNeXt [18] was proposed. To reduce the effect of the class-imbalance problem, different data balancing approaches such as loss weighting and balanced batch sampling were applied. The best results were obtained by a 5-fold cross-validation with an 85.10% accuracy. This method outperformed each of three individual architectures separately. Majtner et al. [19] proposed an automatic prediction system to differentiate between types of skin cancer using an ensemble model of VGG16 and GoogLeNet architectures. Various image augmentation and color normalization methods were performed to balance the dataset and provide better color consistency. Majtner's proposed solution obtained 81.50% accuracy. Lee et al. [20] designed a pipeline, WonDerM, that resamples the pre-processed skin lesion images. Then four fine-tuned neural networks were built based on the data provided by segmentation tasks. To overcome the imbalanced data issue, a balanced sampling method was employed based on weighting

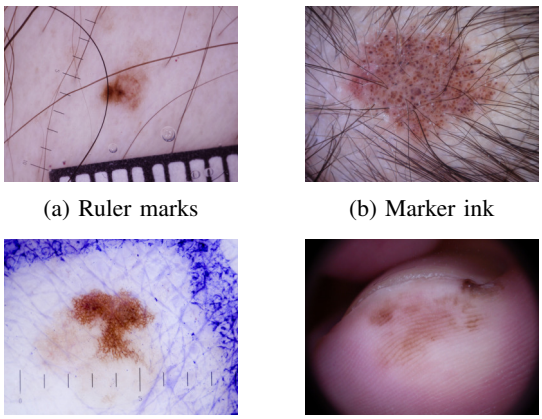
978-1-7281-5341-4/19/$31.00 ©2019 IEEE

prediction probabilities. Lee's proposed ensemble model had higher true positive rates and achieved 78.50% accuracy. Pal et al. [21] proposed an ensemble model based on deep learning for melanoma detection. Three weak CNN classifiers namely, ResNet50, DenseNet121 and MobileNet, combined together to build a stronger classifier. Back-propagation of the weighted loss method was employed to solve the data imbalance issue. Each of the selected pre-trained CNN models were fine-tuned separately, and the final prediction was obtained by averaging across over all classifiers, achieving an accuracy of 77.50%. Finally, a study conducted by Milton [22] proposed a deep ensemble classifier system based on NASNet-5-Large [23], InceptionResNetV2 [24], SENet154, and InceptionV4 for skin lesion classification. Milton's proposed ensemble model obtained 73.00% accuracy and the best result achieved by PNASNet-5-Large architecture with 76.00% accuracy.

Although promising results have been obtained in these works, the possibility of building compact neural networks using depthwise separable CNN for skin cancer classification has not yet been investigated. In this paper, we demonstrate that utilization of depthwise separable CNN with minimal components and without pre-processing steps has the potential to achieve promising results.

B. Motivations and contributions

Accurate detection of skin cancer is a difficult task because of the complicated nature of different types of lesions, i.e. variance in lesion shape, size and textures and also the presence of noise and artifacts. Fig 1 demonstrates some examples of noisy and challenging skin lesion images.



(a) Ruler marks (b) Marker ink

(c) Skin hairs (d) Dark corners

Fig. 1: Typical artifacts contained in dermoscopic images: (a) ruler marks, (b) marker ink, (c) skin hairs and (d) dark corners.

To attempt to compensate for the complicated nature of skin cancer, we investigate the use of separable convolutional networks. Separable convolution initially introduced in [25] and then used in Inception models [24], [26]. It was shown in [27] that employing a depthwise separable network outperformed the MobileNetV2 model on multiple datasets with fewer parameters and faster inference time. Studies by [28] and [29] also showed that data augmentation might further help improve prediction results. The superiority of these previously obtained results motivated us to investigate the use of separable convolutional networks for skin lesion classification tasks.

The contributions of the present study, are twofold. First, for network training, we propose an architecture using separable convolution to extract features from dermoscopic images. Specifically, we investigate several choices for the number of feature maps and pooling window size at each layer. By extracting global and local features from lesions, the proposed CNN architecture may achieve improved results when compared with previous studies. As our second contribution, the advantage of the proposed method could be the detection of skin lesions in the presence of noise in the input data. Compared to conventional methods, where extensive pre-processing methods are used to boost the model performance, the proposed method avoids pre-processing steps or engineered hand-crafted feature descriptors in order to increase the generalization ability. Our proposed architecture achieves an accuracy of 87.24% in classifying the images of ISIC 2018 dataset without any task-specific pre-processing method. To the best of our knowledge, this is the first study toward the implementation of a model using a separable convolution architecture without requiring extensive data pre-processing methods for skin cancer classification.

The rest of this paper is organized as follows: A detailed description of the proposed methodology based on separable convolution neural network is presented in Section II. Results and discussion are reported in Section III. Finally, conclusion and future directions are given in Section IV.

II. Methodology

A. Network architecture

Deep CNNs have achieved state-of-the-art performance in various high-level computer vision tasks in recent years. However, these networks often have millions of parameters and require extensive computational resources. This not suitable for real-time applications with limited processing resources and substantial complexity. In addition, some datasets need extensive pre-processing before feeding into the networks. Depthwise separable CNNs can significantly reduce the number of network parameters and processing time cost with almost the same recognition accuracy. For example, MobileNet [30] based on depthwise separable convolution achieved better performance with only a slight drop in accuracy to 0.9% on ImageNet [31] dataset while being 32 times smaller and 27 times less compute-intensive compared to the VGG16 [32] architecture. Employing depthwise separable convolutions decreases the redundant standard convolution's parameters, leading to the development of more compact networks to overcome the computational complexity. The proposed classification architecture for this experiment is based on a depthwise separable CNN which is illustrated in Fig 3 and consists of six essential components as follows:

- **Depthwise separable convolutional layer:** Depthwise separable convolutions perform the convolution operation, separately and layer by layer. The operation splits into two steps: depthwise and pointwise convolutions, which are illustrated in Fig 2b. Depthwise convolution employs a single filter on each input channel while pointwise convolution uses the outputs from depthwise layer to form a linear combination by 1×1. Pointwise convolution would help to retain spatial information that optimizes the performance of the convolution network.

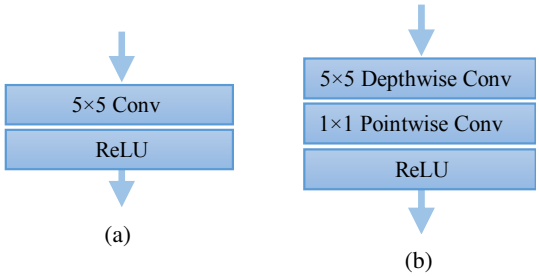


Fig. 2: (a) Standard convolution (b) Depthwise separable convolution including depthwise and pointwise layers.

- **Max Pooling layer:** A max-pooling layer summarizes the feature maps using a kernel size by a constant factor of n×n , i.e., 2 or 3 to capture image information and discard the irrelevant information. This operation takes the maximum value from the convolution window and produces a single output by sub-sampling of the convolutional layer, dramatically reducing the computational cost of the training process. In this way, important texture information is retained and passed to the next convolution layer. Pooling operations help achieve invariance to rotation, scale, distortion or translation, and improve the generalization power of the network.
- **Fully connected layer:** All neurons after convolution and max-pooling operations are fed directly into a 1-dimension vector or fully connected (FC) layer at the end of the network.
- **Dropout layer:** Dropout layer [33] assists in reducing the over-fitting problem by randomly dropping some units to zero at each update during training. Employing the dropout layer improves generalization ability and is especially beneficial when the training data when some classes are imbalanced such as ISIC 2018 dataset in this experiment.
- **Rectified linear unit (ReLU):** ReLU activation function is the most popular nonlinear function that used to accelerate the convergence of stochastic gradient descent for all convolution and FC layers. The ReLU activation function formulated as:

$$f(x) = max(x, 0) \tag{1}$$

- **Softmax function:** The Softmax activation function is used as a classifier or output layer. The Softmax activation function produces a probability distribution that consists of seven neurons corresponding to the 7-class classification problem.

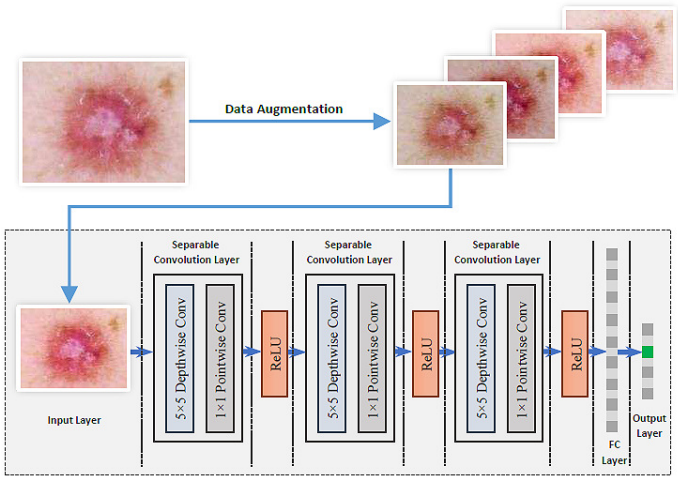


Fig. 3: A schematic diagram of the proposed network architecture using a depthwise separable CNN.

B. Dataset description

This research uses the skin lesion classification dataset of ISIC (International Skin Imaging Collaboration) 2018: Skin Lesion Analysis Towards Melanoma Detection, which is available for the public at [34]. The dataset consists of nearly ten thousand dermoscopic lesion images. The size of each image is 600×450 pixels and acquired from a variety of dermoscopy devices from several institutions. Possible classes in this dataset are: 1) Melanoma (MEL), 2) Melanocytic nevus (NV), 3) Basal cell carcinoma (BCC), 4) Actinic keratosis / Bowen's disease (intraepithelial carcinoma) (AKIEC), 5) Benign keratosis (solar lentigo / seborrheic keratosis / lichen planus-like keratosis) (BKL), 6) Dermatofibroma (DF), 7) Vascular lesion (VASC).

C. Data pre-processing

Several works in the literature have performed skin lesion classification using extensive pre-processing steps including color space transformation, contrast enhancement, illumination correction, and noise reduction methods [35]–[37]. The purpose is to improve the quality of input images and remove artifacts such as hair, ruler markers, reflections, bubbles and shadows. However, training a network with elaborate pre-processing steps may degrade the generalization ability. Hence, the proposed method avoids using extensive pre-processing steps and also hand-crafted feature descriptors to support better generalization ability and make our method robust to noise and artifacts. In this study, we employed only two standard pre-processing steps, which are common in training deep learning models. First, for image normalization, we rescaled the intensity values of the pixels from [0,1] to have a uniform distribution. Another pre-processing method is data standardization with a zero mean and a standard deviation equal to one to remove bias from the features.

D. Data augmentation

The class distribution of ISIC 2018 dataset is highly imbalanced, i.e. 68%, 11% and 10% for the NV, BKL and MEL classes , respectively. Different data augmentation methods for minority classes such as horizontal and vertical flips, random filter, and rotation are employed to balance the minority classes of the dataset and avoid the negative effect of class imbalances. Some examples of data augmentation step are illustrated in Fig 4.

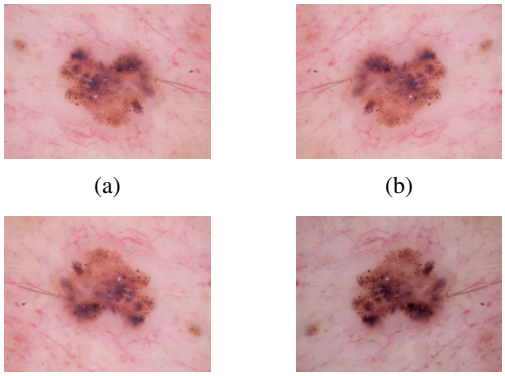


Fig. 4: Examples of the augmented dataset. (a) Original image, (b) Horizontal flip (c) Rotation, horizontal and vertical flip, (d) Random filter and vertical flip.

E. Evaluation Metrics

To quantify the performance of the proposed method, three evaluation metrics, namely; accuracy, sensitivity and specificity, are used. Given the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN), the measures are mathematically expressed as follows:

$$Accuracy(ACC) = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (2)$$

$$Sensitivity(Sens) = \frac{TP}{TP + FN} \times 100 \quad (3)$$

$$Specificity(Spec) = \frac{TN}{TN + FP} \times 100 \quad (4)$$

III. Experiment and Results

A. Experimental Setup

Some of the experimental setups are as follows: $\beta 1$, $\beta 2$ and learning rate for the Adam optimizer were set to 0.6, 0.995 and 0.0001, respectively. To avoid overfitting, the dropout layer has been included with a ratio of 0.5 to the fully connected layer. The number of epochs and batch size were set to 50 and 8, respectively. Rectified linear unit (ReLU) is employed for training the model. We used the original image size of 600×450 pixels. Training and testing process of the proposed architecture is performed using NVIDIA GTX 1080 TI with 11 GB graphical processing unit (GPU) memory, and the operating system is Windows with an Intel(R) Core(TM) i7-8700K 3.7 GHz processors with 32 GB RAM. The network architectures are implemented in Python using the Keras package with Tensorflow backend.

B. Results and discussion

The presented architecture relies on depthwise and pointwise convolutional blocks to decrease the dimensionality of the input images and at the same time enhance the extraction of representational features. Details of the experiment results are summarized in Table I.

Various architectures with different depth can be designed by employing a different combination of hyper-parameter settings such as the number of convolution layers, input image dimension, kernel size and optimizers. However, implementing such exhaustive experiments is highly time-consuming and limited due to the available hardware and GPU memory. Hence, we considered the best possible architectures for this task. We investigate the effectiveness of different size of separable convolution kernel window, i.e. 3×3, 5×5, 7×7, 9×9 in each convolution layer on the performance of the model. The impact of the number of feature maps per separable convolution layers (32, 64, 128, 256, 512) also examined. Max-pooling layers with a filter size of 2×2 pixels and a stride of 1×1 pixels are used for the proposed architecture, and the fully connected hidden layer is composed of 512 neurons. By trial and error, the best result achieved by model 6, as highlighted in Table I, with the structure of 4 separable convolution layers with a kernel size of 5×5, one fully-connected layer, one dropout layer, and one classification layer with a Softmax activation function. We found that employing convolution layers with a kernel size of 5×5 followed by smaller max pooling layers with a kernel size of 2×2 results in a more stable learning process and hence increase the performance of the method. We also modified the architecture with both smaller and larger kernel size than 5×5. It observed that employing 3 separable layers with larger kernel size such as 7×7 and 9×9 gave a poor performance as accuracy decreased in models 4 and 5, while applying smaller kernel size in the architecture results in a better detection rate as demonstrated in models 2, 6 and 9; thus, any increase in both the number of feature maps and kernel size cause the information loss and poor performance of the model. Furthermore, implementing a deeper network with more separable convolution layers dramatically increases the number of parameters and model size leading to over-fitting problem and the computational time complexity constraint. Also, it should be emphasized that models 5,7 and 11 achieved 100% specificity, and models 5 and 9 achieved 97.10% and 97.14% sensitivity which are higher than the presented architecture with 95.94% sensitivity and 98.47% specificity.

C. Comparison with related works

Referring to Table II, the methods in [12], [15], [19]–[22] give accuracies of 87.20%, 85.10%, 81.50%, 78.50%, 77.50% and 76.00% respectively, whereas, the results obtained using

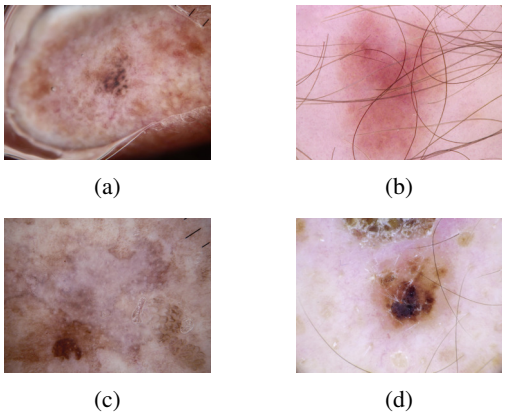
TABLE I: Details of implemented depthwise CNN architectures for skin cancer classification on ISIC 2018 dataset.

Model	# separable layers	kernel size: separable layers	kernel size: pooling layers	# feature maps	ACC (%)	Sens (%)	Spec (%)
1	3	3,3	2,2	32, 64, 128	86.47	90.54	99.2
2	3	5,5	2,2	32, 64, 128	87.2	94.11	99.17
3	3	7,7	2,2	32, 64, 128	81.17	94.59	99.23
4	3	9,9	2,2	32, 64, 128	82.94	94.52	97.79
5	4	3,3	2.2	32, 64, 128, 256	84.74	97.1	100
6	4	5,5	2,2	32, 64, 128, 256	**87.24**	95.94	98.47
7	4	7,7	2,2	32, 64, 128, 256	85.01	94.66	100
8	4	9,9	2,2	32, 64, 128, 256	86.16	89.87	98.46
9	5	3,3	2,2	32, 64, 128, 256, 512	85.55	97.14	97.77
10	5	5,5	2,2	32, 64, 128, 256, 512	81.86	92.3	99.23
11	5	7,7	2,2	32, 64, 128, 256, 512	86.55	93.58	100
12	5	9,9	2,2	32, 64, 128, 256, 512	83.28	91.89	98.46

the presented framework obtained an accuracy of 87.24%, which is 0.04% higher than the best method and 11.24% higher than the worst accuracy from previous methods applied to the same dataset. Referring to Table II, we can see that the self-designed model presented here, achieved a favorable result in terms of accuracy compared with the individual or ensemble models previously described in the literature, i.e. SENet, ResNeXt, DenseNet, GoogLeNet, ResNet50 and MobileNet. From the obtained results, we conclude that our model can extract distinguishing features efficiently and achieve a better result in terms of accuracy. Additionally, the method presented here has better generalization ability to the noise compared to those with extensive pre-processing steps [12], [22] or external data sources [15].

TABLE II: Comparative analysis of the proposed approach with other CNN architecture presented in the literature. .

Method	ACC (%)
Kitada et al. [12]	87.2
Gessert et al. [15]	85.1
Majtner et al. [19]	81.5
Lee et al. [20]	78.5
Pal et al. [21]	77.5
Milton [22]	76
Proposed model	**87.24**



(a) (b) (c) (d)

Fig. 5: Examples of misclassified lesions.

Although the approach presented here shows satisfying performance, it has also limitations on classifying more challenging instances with vague, low contrast boundaries, and also the presence of artifacts such as hair, reflections and bubbles as illustrated in Fig 5. The presence of such artifacts harms the performance of the model and can adversely affect the detection rate.

IV. Conclusion

This paper studies the use of a depthwise separable CNN for skin cancer detection. The results reach an accuracy rate of 87.24%. These results show that the approach presented in this study is still optimal in terms of accuracy compared to other methods. The model, achieved comparable performance to that of standard convolution and also ensemble deep CNN while required no pre-processing techniques specific to skin lesion classification task such as noise reduction without compromising accuracy. We conclude that even though the presented architecture is compact due to the separation operation; the network shows good performance and yields satisfying results. Future work should consider the collection of more datasets to test the robustness of the proposed separable CNN model. Also, to develop more efficient and accurate CAD systems, a detailed study is needed to determine the relationship between parameters in designing separable CNN architecture.

References

[1] "Key Statistics for Melanoma." [Online]. Available: https://www.cancer.org/cancer/melanoma-skin-cancer/about/key-statistics.html

[2] S. H. Kassani, P. H. Kassani, M. J. Wesolowski, K. A. Schneider, and R. Deters, "Breast cancer diagnosis with transfer learning and global pooling," *arXiv preprint arXiv:1909.11839*, 2019.

[3] S. H. Kassani, P. H. Kassani, M. J. Wesolowski, K. A. Schneider, R. Deters *et al.*, "A hybrid deep learning architecture for leukemic b-lymphoblast classification," *arXiv preprint arXiv:1909.11866*, 2019.

[4] S. H. Kassani, P. H. Kassani, M. J. Wesolowski, K. A. Schneider, and R. Deters, "Classification of histopathological biopsy images using ensemble of deep learning networks," *arXiv preprint arXiv:1909.11870*, 2019.

[5] R. Rasti, M. Teshnehlab, and S. L. Phung, "Breast cancer diagnosis in dce-mri using mixture ensemble of convolutional neural networks," *Pattern Recognition*, vol. 72, pp. 381–390, 2017.

[6] Y. Ariji, Y. Yanashita, S. Kutsuna, C. Muramatsu, M. Fukuda, Y. Kise, M. Nozawa, C. Kuwada, H. Fujita, A. Katsumata *et al.*, "Automatic detection and classification of radiolucent lesions in the mandible on panoramic radiographs using a deep learning object detection technique," *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, 2019.

[7] V. Jahmunah, S. L. Oh, J. K. E. Wei, E. J. Ciaccio, K. Chua, T. R. San, and U. R. Acharya, "Computer-aided diagnosis of congestive heart failure using ecg signals–a review," *Physica Medica*, vol. 62, pp. 95–104, 2019.

[8] M. A. Al-Masni, M. A. Al-antari, M.-T. Choi, S.-M. Han, and T.-S. Kim, "Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks," *Computer methods and programs in biomedicine*, vol. 162, pp. 221–231, 2018.

[9] N. Nida, A. Irtaza, A. Javed, M. H. Yousaf, and M. T. Mahmood, "Melanoma lesion detection and segmentation using deep region based convolutional neural network and fuzzy c-means clustering," *International journal of medical informatics*, vol. 124, pp. 37–48, 2019.

[10] S. Guo and Z. Yang, "Multi-channel-resnet: an integration framework towards skin lesion analysis," *Informatics in Medicine Unlocked*, vol. 12, pp. 67–74, 2018.

[11] N. Ibtehaz and M. S. Rahman, "Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation," *arXiv preprint arXiv:1902.04049*, 2019.

[12] S. Kitada and H. Iyatomi, "Skin lesion classification with ensemble of squeeze-and-excitation networks and semi-supervised learning," *arXiv preprint arXiv:1809.02568*, 2018.

[13] Y. Tokozume, Y. Ushiku, and T. Harada, "Between-class learning for image classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5486–5494.

[14] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," *arXiv preprint arXiv:1708.04896*, 2017.

[15] N. Gessert, T. Sentker, F. Madesta, R. Schmitz, H. Kniep, I. Baltruschat, R. Werner, and A. Schlaefer, "Skin lesion diagnosis using ensembles, unscaled multi-crop evaluation and loss weighting," *arXiv preprint arXiv:1808.01694*, 2018.

[16] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[17] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[18] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1492–1500.

[19] T. Majtner, B. Bajić, S. Yildirim, J. Y. Hardeberg, J. Lindblad, and N. Sladoje, "Ensemble of convolutional neural networks for dermoscopic images classification," *arXiv preprint arXiv:1808.05071*, 2018.

[20] Y. C. Lee, S.-H. Jung, and H.-H. Won, "Wonderm: Skin lesion classification with fine-tuned neural networks," *arXiv preprint arXiv:1808.03426*, 2018.

[21] A. Pal, S. Ray, and U. Garain, "Skin disease identification from dermoscopy images using deep convolutional neural network," *arXiv preprint arXiv:1807.09163*, 2018.

[22] M. A. A. Milton, "Automated skin lesion classification using ensemble of deep neural networks in isic 2018: Skin lesion analysis towards melanoma detection challenge," *arXiv preprint arXiv:1901.10802*, 2019.

[23] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8697–8710.

[24] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[25] L. Sifre and S. Mallat, "Rigid-motion scattering for image classification," *Ph. D. dissertation*, 2014.

[26] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[27] Y. Xiong, H. J. Kim, and V. Hedau, "Antnets: Mobile convolutional neural networks for resource efficient image classification," *arXiv preprint arXiv:1904.03775*, 2019.

[28] S. H. Kassani and P. H. Kassani, "A comparative study of deep learning architectures on melanoma detection," *Tissue and Cell*, vol. 58, pp. 76–83, 2019.

[29] Z. Shi, M. Liu, Q. Cao, H. Ren, and T. Luo, "A data augmentation method based on cycle-consistent adversarial networks for fluorescence encoded microsphere image analysis," *Signal Processing*, vol. 161, pp. 195–202, 2019.

[30] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[32] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[33] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[34] "ISIC 2018." [Online]. Available: https://challenge2018.isic-archive.com/

[35] N. Hameed, A. M. Shabut, M. K. Ghosh, and M. Hossain, "Multi-class multi-level classification algorithm for skin lesions classification using machine learning techniques," *Expert Systems with Applications*, p. 112961, 2019.

[36] S. Chatterjee, D. Dey, S. Munshi, and S. Gorai, "Extraction of features from cross correlation in space and frequency domains for classification of skin lesions," *Biomedical Signal Processing and Control*, vol. 53, p. 101581, 2019.

[37] A. Mahbod, G. Schaefer, I. Ellinger, R. Ecker, A. Pitiot, and C. Wang, "Fusing fine-tuned deep features for skin lesion classification," *Computerized Medical Imaging and Graphics*, vol. 71, pp. 19–29, 2019.