Web-based Melanoma Detection

SangHyuk Kim[1][0009−0005−1323−306X], Edward Gaibor[2][0009−0009−0400−7529], and Daniel Haehn[3][0000−0001−9144−3461]

[1] University of Massachusetts Boston, sanghyuk.kim001@umb.edu
[2] University of Massachusetts Boston, edward.gaibor001@umb.edu
[3] University of Massachusetts Boston, daniel.haehn@umb.edu

Abstract. Melanoma is the most aggressive form of skin cancer, and early detection can significantly increase survival rates and prevent cancer spread. However, developing reliable automated detection techniques is difficult due to the lack of standardized datasets and evaluation methods. This study introduces a unified melanoma classification approach that supports 54 combinations of 11 datasets and 24 state-of-the-art deep learning architectures. It enables a fair comparison of 1,296 experiments and results in a lightweight model deployable to the web-based MeshNet architecture named Mela-D. This approach can run up to 33x faster by reducing parameters 24x to yield an analogous 88.8% accuracy comparable with ResNet50 on previously unseen images. This allows efficient and accurate melanoma detection in real-world settings that can run on consumer-level hardware.

Keywords: Skin cancer classification · Medical imaging application · Supervised learning.

1 Introduction

Melanoma is a deadly skin cancer, but early detection and treatment can increase survival rates. Convolutional neural networks (CNNs) have improved melanoma detection rates by utilizing open melanoma datasets, surpassing feature-based classifiers [2]. However, developing robust and practically deployable CNN-based melanoma classifiers has several challenges. Firstly, a classifier must use diverse and representative training datasets to deal with the wide variability in melanoma appearance (Fig. 1). However, existing methods [1,11,20,23,22] often use limited training data with non-standardized preprocessing protocols, hindering the generalizable and reproducible classifiers. Secondly, assessing models is often inconsistent across melanoma studies, with different datasets, preprocessing methods, and performance metrics being used. This inconsistency makes it challenging to compare each approach, as shown in (Fig. 2). Lastly, it is important to note that when classification models are generated from incompatible frameworks or undergo different preprocessing, they perform poorly or even become difficult to deploy (Fig. 2). This makes the models unusable for web-based applications, which can limit dermatologists' access to these models in real-world

clinical settings. This paper proposes a comprehensive framework for melanoma classification that enables robust evaluation and efficient web deployment (Fig. 3).

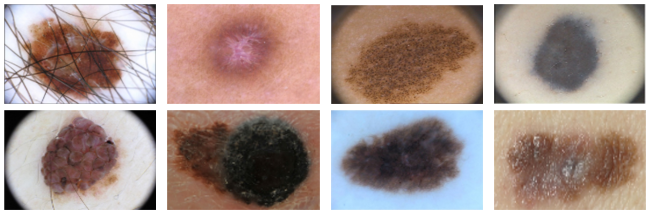

Fig. 1: Examples of melanoma images from different datasets illustrate the diversity in appearance and challenges for automated detection. Variations in size, shape, color, and texture across datasets highlight the need for a robust and generalizable melanoma detection approach.

Our key contributions are as follows: We curate a diverse set of 11 publicly available melanoma datasets to train and evaluate our models. This allows us to train the classifiers with various data distributions and image characteristics. We then conduct a large-scale study, training 24 state-of-the-art CNN architectures on each of the 54 dataset combinations, resulting in a total of 1,296 experiments.

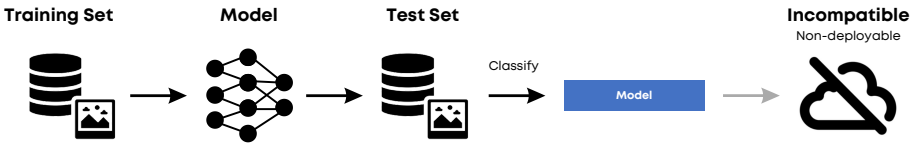

Fig. 2: The typical workflow of existing melanoma classifiers is non-deployable for web-based applications as each model is generated from incompatible frameworks or undergoes different preprocessing. This limits practical utility for end-users such as dermatologists.

This enables us to identify the best-performing models and to assess the impact of dataset composition on classification performance. We then introduce Mela-D, a lightweight CNN architecture optimized for web deployment. Based on a MeshNet [9] model, Mela-D utilizes dilated convolutions [30] to capture multi-scale features. Mela-D achieves a parameter reduction of up to 24x compared to conventional CNN architectures, making it suitable for real-time inference in web-based applications. We integrate Mela-D into an open-source web-based platform that allows users to upload dermatoscopic images and quickly receive real-time melanoma classification results.

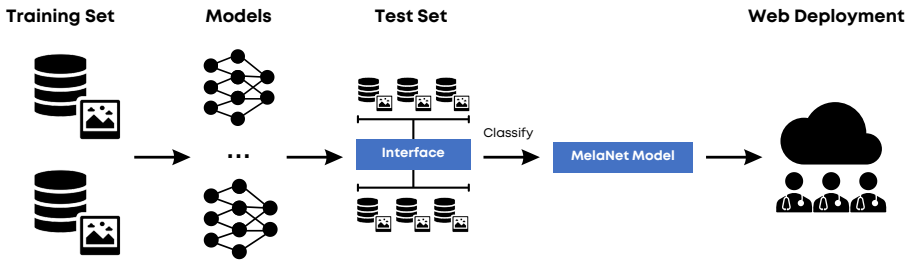

Fig. 3: The figure is an overview of our proposed melanoma classification framework, which is deployable on the web through a user-friendly interface. It enables end-users to easily classify melanoma images.

2 Methods

2.1 Data Interface

The melanoma datasets [29,12,5,4,6,24,19,17,21,10,8] consist of JPEG and PNG images and metadata containing binary labels (benign/malignant). The image size varies between 147 x 147 and 3096 x 3096, and the directory structure varies depending on the collecting institutions. While most datasets provide ground truth in the form of comma-separated values (CSV), some datasets have folders divided for benign/malignant cases serving as labels. Our framework is capable of converting a wide range of melanoma datasets [29,12,5,4,6,24,19,17,21,10,8] and freely assembling them in any combination to be preprocessed and trained in CNNs. (Fig. 4).

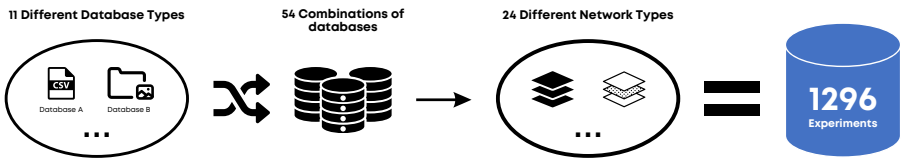

Fig. 4: The 11 melanoma databases containing images and labels are combined into 54 different sets. These sets are trained and evaluated with 24 network architectures, yielding 1,296 experiments.

2.2 Classification Methods

We explore various CNN architectures as base networks to identify the best-performing melanoma classifier. The list of all CNNs tested are [16]: DenseNet121 / DenseNet169 / DenseNet201, [28]: EfficientNet B0 to B7 [27]: InceptionV3,

[15]: MobileNet, [25]: MobileNetV2, [13]: ResNet50 / ResNet101 / ResNet152, [14]: ResNet50V2 / ResNet101V2 / ResNet152V2, [26]: VGG16/VGG19, [3]: Xception. We use the Adam optimizer [18] with 0.0001 learning rate, 32 batch size, 20 epochs, and categorical cross-entropy loss training from ImageNet pre-trained weights for all models except Mela-D. Each network's binary softmax activation function predicts the benign/malignant class probabilities. We resize every image to 150x150 pixels and apply data augmentation (rotation, zoom, cropping, flipping) to balance the benign/malignant ratio to 50:50.

2.3 Mela-D and Web Deployment

MeshNet architecture [9] employs a compact CNN model using dilation convolution [30] to flexibly control the size of the receptive field of the input neuron. In this regard, the model can integrate multiple contexts without increasing the number of layers and parameters while maintaining comparable performance to the existing CNNs. Dilated convolution is represented as:

$$(F *_l k)(\mathbf{p}) = \sum_{\mathbf{s}+l\mathbf{t}=p} F(\mathbf{s})\, k(\mathbf{t}) \tag{1}$$

where l is dilation factor and $*_l$ is l-dilated convolution. Based on MeshNet [9], we build a Mela-D architecture (Fig. 5) to reduce the number of parameters by 24x compared to the existing network by flexibly controlling the receptive field size. All the layers independently perform inference and do not lose information in feed-forward.

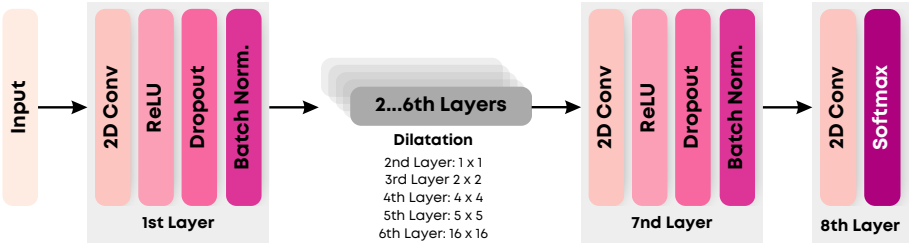

Fig. 5: Mela-D architecture uses Dilated convolutions [30] with increasing dilation factors in each layer to achieve independent computation during inference with lower computational cost.

3 Experiments and Results

We conduct experiments using Python 3.9, Keras 2.5.0rc0, TensorFlow 2.5.0, and high-performance computing resources (AMD EPYC 7742 CPUs, Nvidia DGX-A100 GPUs, 2064GB RAM). Runtime benchmarks are performed using Google

Chrome Performance Monitor on a Windows 11 laptop (Intel Core i5-1135G7 @ 2.40GHz, Intel Iris Xe Graphics, 8GB RAM) with WebGL TensorFlow.js backend.

3.1 Performance

We comprehensively evaluate 1,296 experiments on five test sets [29,12,5,4,17] using various performance metrics such as precision, sensitivity, specificity, F1-score, and accuracy. The models are ranked based on precision, the primary metric used in the ISIC challenge [12]. We then select the top three models for each test set based on this ranking metric. As can be seen in Table 1, ResNet152 [13], ResNet50 [13], and DenseNet121 [16] are consistently among the best-performing classifiers.

Table 1: Top-3 performing models per combination of dataset and classifier. The performance is calculated based on precision, specificity, sensitivity, and accuracy. The results are in precision descending order. The best-performing metric per test set is in **bold**.

Testset	Classifier	Trainsets	Precision	Specificity	Sensitivity	Accuracy
HAM10000 [29]	ResNet152 [13]	a+b+c+d+e+g	**99.3%**	**100%**	**70.6%**	**98.6%**
	VGG16 [26]	a+b+c+d+e+f+g	99.1%	**100%**	62.7%	98.3%
	EfficientNetB2 [28]	a+b+c+d+e	99.0%	**100%**	60.8%	98.2%
ISIC2016 [12]	ResNet152 [13]	a+b+c+d+e+g	**96.3%**	**100%**	68.0%	93.7%
	ResNet152 [13]	a+b+c+d+e	96.2%	99.7%	**74.7%**	**94.7%**
	ResNet50 [13]	a+b+c+d+h+i	96.0%	99.7%	73.3%	94.4%
ISIC2017 [5]	ResNet152[13]	d	**91.3%**	99.2%	49.6%	**89.5%**
	VGG19[26]	a+b+c+d+e+h+i	91.3%	98.8%	56.4%	90.5%
	EfficientNetB0 [26]	a+f	90.4%	**100%**	2.0%	80.8%
ISIC2018 [4]	ResNet50[13]	c	**76.1%**	**97.3%**	32.2%	89.9%
	DenseNet201[16]	a+b+c+d+e+f+h	75.9%	96.2%	42.7%	**90.1%**
	DenseNet121[16]	e	75.9%	99.6%	5.8%	89.0%
7-point criteria [17]	ResNet50[13]	e	**87.3%**	**100%**	0.9%	74.7%
	ResNet50[13]	a+b+c+h+g	33.7%	98.0%	**81.5%**	74.7%
	DenseNet121[16]	a+f+h	82.2%	99.0%	18.8%	**78.5%**

a:ISIC16 [12] b:ISIC2017 [5] c:ISIC2018 [4] d:ISIC2019 [5] e:ISIC2020 [24]
f:7-point criteria [17] g:PH2 [19] h:PAD_UFES_20[21] i: MEDNODE [10]
j: Kaggle [8]

However, we can observe that the best-performing models have slower runtimes on the browser than the web-deployable Mela-D (Table 2).

Table 2: We measure the runtime of the best-performing on the browser. Each classifier has the highest precision on each test set. We conduct three trials to measure the runtimes. The fastest runtime is in **bold**. Our observations show that our proposed model, Mela-D, is the fastest classifier among all the ones we test.

Classifier	Trainsets	1st trial	2nd trial	3rd trial	Average
Proposed method	a+b+c+d+e+g+f+i+j+k	**682 ms**	**654 ms**	**621 ms**	**652.3 ± 30.5 ms**
ResNet152 [13]	a+b+c+d+e+g	9125 ms	6129 ms	7177 ms	7477 ± 1520.4 ms
DenseNet121 [16]	a+b+c+d+i+j	22204 ms	21987 ms	20683 ms	21624.7 ± 822.7 ms
ResNet50 [13]	c	1791 ms	1438 ms	1663 ms	1630.7 ± 178.7 ms
ResNet152 [13]	d	12065 ms	9328 ms	12017 ms	11136.7 ± 1566.5 ms
ResNet50 [13]	e	7264 ms	7546 ms	7669 ms	7493 ± 207.6 ms

a:ISIC16 [12] b:ISIC2017 [5] c:ISIC2018 [4] d:ISIC2019 [6] e:ISIC2020 [24] f:7-point criteria [17] g:PH2 [19] h:PAD_UFES_20 [21] i:MEDNODE [10] j:Kaggle [8]

To investigate this aspect, we deploy the models on the web and perform three execution trials, calculating the average runtime in seconds. The results in (Fig. 6) show that Mela-D outperforms the other models, demonstrating a minimum of twice and a maximum of 33 times faster execution on the browser.

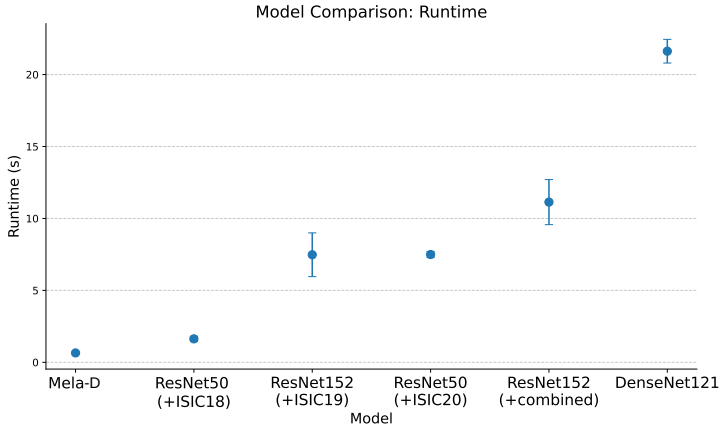

Fig. 6: Comparison of average execution time (in seconds) for Mela-D and the top-performing models from each test set. Mela-D demonstrates 2 to 33 times faster runtimes than the other models. Tested with Google Chrome on Intel Core i5-1135G7 @ 2.40GHz, Intel Iris Xe Graphics, 8GB RAM, with WebGL TensorFlow.js backend.

To evaluate the scalability of Mela-D when considering running it on consumer devices, we additionally analyze the ratio between performance and run-

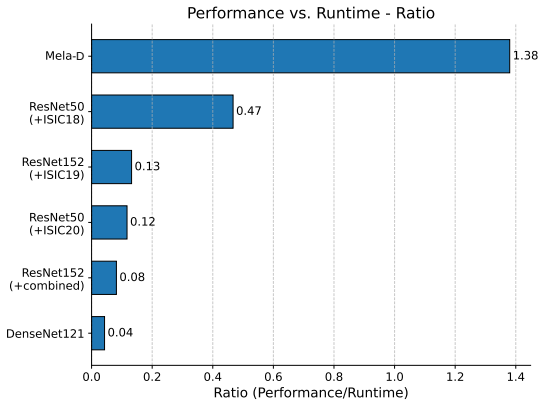

Fig. 7: Mela-D achieves the highest ratio, indicating a high balance between classification accuracy and runtime speed. Thus, Mela-D is well-suited for efficient web-based deployment and real-time inference.

time. We assess the performance (precision) vs. runtime ratio of Mela-D and the top-performing models from each test set to verify the efficiency of performance versus runtime. The formula divides precision by its average runtime in seconds. The results are shown in (Table 2). The analysis reveals that Mela-D achieves the highest performance-time ratio among the best-performing models with accuracy 88.8% on ISIC18 [4] benchmark and 652 ms runtime on the web. At the same time, ResNet50 [13] records accuracy 89.9% on ISIC18[4] and 1630 ms runtime on the web, indicating Mela-D's superiority in performance versus computational efficiency (Fig. 7).

3.2 Web Deployment

We deploy Mela-D to the web through integration with the open-source platform at https://boostlet.org, allowing client-side image processing via custom browser bookmarklets that inject JavaScript functionality when visiting websites. We convert Mela-D into a TensorflowJS layers model to exploit its lightweight nature through a simple drag-and-drop interface. When visiting a website containing skin images such as DermNetNZ.org [7], users can click the bookmarklets, drag an image onto the page, and receive a melanoma classification result via a front-end alert (see Fig. 8 and supplementary video). We also have a standalone web application where users can test local images at https://mpsych.github.io/melanoma/ as shown in Fig. 9.

4 Conclusions

We have introduced a melanoma detection framework with a lightweight machine learning model, Mela-D, to leverage existing databases to build a robust

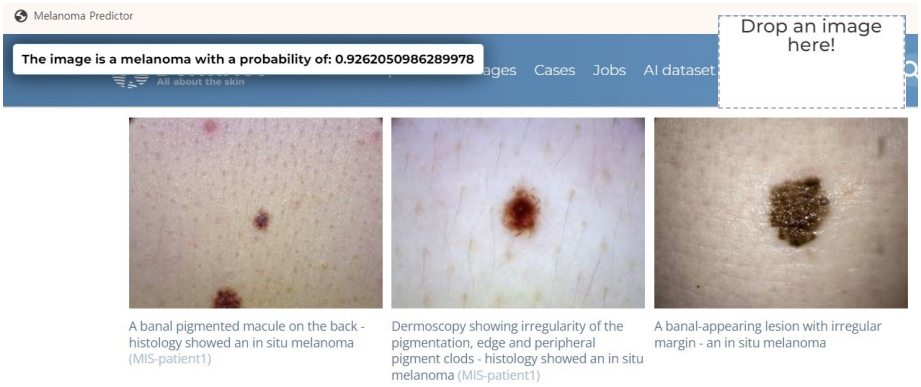

Fig. 8: Mela-D, deployed on Boostlet.js, a web-based image processing plugin platform, performs melanoma prediction on images on the web with a drag-and-drop interface.

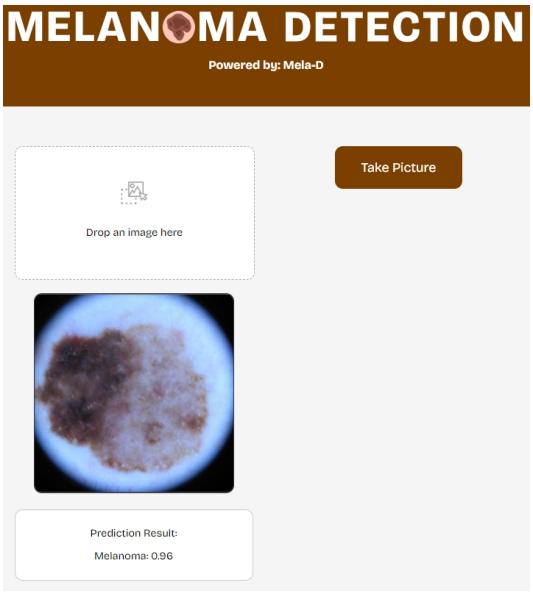

Fig. 9: Users can determine skin cancer from any image they capture through a standalone web application.

and compact classifier for the web. We verify its efficiency by comparing the performance versus runtime ratio with existing CNNs and deploying the model on the web. In the future, we plan to build an affordable skin change tracker based on this framework. We release our work and results as open source at https://mpsych.org/melanoma.

References

1. Adegun, A.A., Viriri, S.: Deep learning-based system for automatic melanoma detection. IEEE Access **8**, 7160–7172 (2019)
2. Bhatt, H., Shah, V., Shah, K., Shah, R., Shah, M.: State-of-the-art machine learning techniques for melanoma skin cancer detection and classification: a comprehensive review. Intelligent Medicine **3**(03), 180–190 (2023)
3. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1251–1258 (2017)
4. Codella, N., Rotemberg, V., Tschandl, P., Celebi, M.E., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M., et al.: Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). arXiv preprint arXiv:1902.03368 (2019)
5. 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)
6. Combalia, M., Codella, N.C., Rotemberg, V., Helba, B., Vilaplana, V., Reiter, O., Carrera, C., Barreiro, A., Halpern, A.C., Puig, S., et al.: Bcn20000: Dermoscopic lesions in the wild. arXiv preprint arXiv:1908.02288 (2019)
7. DermNet: https://dermnetnz.org/
8. FANCONI, C.: https://www.kaggle.com/datasets/fanconic/skin-cancer-malignant-vs-benign
9. Fedorov, A., Johnson, J., Damaraju, E., Ozerin, A., Calhoun, V., Plis, S.: End-to-end learning of brain tissue segmentation from imperfect labeling. In: 2017 International Joint Conference on Neural Networks (IJCNN). pp. 3785–3792. IEEE (2017)
10. Giotis, I., Molders, N., Land, S., Biehl, M., Jonkman, M.F., Petkov, N.: Med-node: A computer-assisted melanoma diagnosis system using non-dermoscopic images. Expert systems with applications **42**(19), 6578–6585 (2015)
11. Gulati, S., Bhogal, R.K.: Detection of malignant melanoma using deep learning. In: International conference on advances in computing and data sciences. pp. 312–325. Springer (2019)
12. 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)
13. 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)
14. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14. pp. 630–645. Springer (2016)
15. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 (2017)

16. 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)
17. Kawahara, J., Daneshvar, S., Argenziano, G., Hamarneh, G.: Seven-point checklist and skin lesion classification using multitask multimodal neural nets. IEEE journal of biomedical and health informatics **23**(2), 538–546 (2018)
18. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
19. Mendonça, T., Ferreira, P.M., Marques, J.S., Marcal, A.R., Rozeira, J.: Ph 2-a dermoscopic image database for research and benchmarking. In: 2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC). pp. 5437–5440. IEEE (2013)
20. Nida, N., Irtaza, A., Javed, A., Yousaf, M.H., Mahmood, M.T.: Melanoma lesion detection and segmentation using deep region based convolutional neural network and fuzzy c-means clustering. International journal of medical informatics **124**, 37–48 (2019)
21. Pacheco, A.G., Lima, G.R., Salomao, A.S., Krohling, B., Biral, I.P., de Angelo, G.G., Alves Jr, F.C., Esgario, J.G., Simora, A.C., Castro, P.B., et al.: Pad-ufes-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones. Data in brief **32**, 106221 (2020)
22. Patel, T.: Tirth27/skin-cancer-classification-using-deep-learning, https://github.com/Tirth27/Skin-Cancer-Classification-using-Deep-Learning
23. Rodrigues, D.d.A., Ivo, R.F., Satapathy, S.C., Wang, S., Hemanth, J., Reboucas Filho, P.P.: A new approach for classification skin lesion based on transfer learning, deep learning, and iot system. Pattern Recognition Letters **136**, 8–15 (2020)
24. Rotemberg, V., Kurtansky, N., Betz-Stablein, B., Caffery, L., Chousakos, E., Codella, N., Combalia, M., Dusza, S., Guitera, P., Gutman, D., et al.: A patient-centric dataset of images and metadata for identifying melanomas using clinical context. Scientific data **8**(1), 34 (2021)
25. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4510–4520 (2018)
26. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
27. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2818–2826 (2016)
28. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. pp. 6105–6114. PMLR (2019)
29. 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)
30. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122 (2015)