



HAL
open science

Hierarchical approach for the classification of multi-class skin lesions based on deep convolutional neural networks

Samia Benyahia, Boudjelal Meftah, Olivier Lézoray

► To cite this version:

Samia Benyahia, Boudjelal Meftah, Olivier Lézoray. Hierarchical approach for the classification of multi-class skin lesions based on deep convolutional neural networks. ICPRAI, Jun 2022, Paris, France. 10.1007/978-3-031-09282-4_12 . hal-03688422

HAL Id: hal-03688422

<https://hal.science/hal-03688422>

Submitted on 4 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Hierarchical approach for the classification of multi-class skin lesions based on deep convolutional neural networks

Samia Benyahia¹, Boudjelal Meftah²[0000-0001-5772-1878], and Lézoray Olivier³[0000-0003-0540-543X]

- ¹ Department of Computer Science, Faculty of Exact Sciences, University of Mascara
Mascara, Algeria
`benyahia.samia@univ-mascara.dz`
- ² LRSBG Laboratory, University of Mascara
Mascara, Algeria
- ³ Normandie Université, UNICAEN, ENSICAEN, CNRS, GREYC
Caen, France

Abstract. Skin lesion is one of the most critical challenges nowadays due to the difficulty of distinguishing a benign lesion from a malignant one. Melanoma represents a malignant melanocytic type of cancer among the most dangerous ones. In contrast, basal cell carcinoma and squamous cell carcinoma represent no malignant melanocytic types of cancer that threaten many human lives. Fortunately, there is some possibility of a cure if it is early detected and well treated. Currently, dermatologists use a hierarchical visual categorization of the lesion or skin biopsy for the diagnostic of skin lesion types. However, computer-aided detection methods can be more accurate, faster, and less expensive than human based techniques. We propose to combine both strategies to develop an efficient skin lesion classification models: the hierarchical organization accredited by dermatologists and a deep learning architecture. In this work, we propose a new hierarchical model for detecting various types of skin lesions based on the combination of several models of convolutional neural networks, where each model is specialized to some types of skin lesion according to the taxonomy. The obtained results highlight the benefits of addressing the classification of different skin lesions with CNNs in such a hierarchically structured way.

Keywords: Classification · Convolutional Neural Networks · Hierarchical Diagnosis · Skin Lesion Diagnosis

1 Introduction

Human skin represents the most significant organ covering the body. The skin acts as a huge sensor packed with nerves to keep the brain in touch with the outside world. At the same time, it acts as a waterproof, helps regulating body temperature and damaging sunlight, additionally allowing us free movement,

and protecting us from microbes. Human skin is made up of three layers. The epidermis that represents the outermost layer of the skin and consists mainly of cells called keratinocytes. The dermis is beneath the epidermis and consists of tough connective tissue and hair follicles. Then, the hypodermis, the deeper subcutaneous tissue, consists of fat and connective tissue. The melanin pigmentation of existing mesodermal cells can absorb UV radiations and cause cancers capable of spreading to other organs. Skin cancer is considered as one of the most diagnosed cancer in the world, and its diagnosis in a precise time is an essential health requirement. However, every misdiagnosis of skin cancer can lead to the death of a patient. Dermatologists experts declare that if the cancer is early diagnosed and followed by an appropriate treatment, they can avoid the patient's death and therefore improve global healthcare. Due to the difficulty of analyzing skin images and detecting cancer with traditional medical techniques such as the ABCD rule or the the 7 point list, various methods have been proposed.

In recent years, we have witnessed an extensive use of Convolution Neural Networks (CNN), that have outperformed others conventional machine learning approaches. CNNs reduce the need of manually handcrafting features by learning highly discriminative features while being trained in an end-to-end supervised way. Several CNN models have successfully been used for the identification of skin cancer lesions with higher efficiency in a reasonable time. We review some of them in the sequel. Sara Atito et al. [1] proposed a deep learning method for skin lesion classification by fusing and fine-tuning an ensemble of three pre-trained deep learning architectures (Xception, Inception ResNet-V2, and NasNetLarge) using the ISIC2019 database and achieved an accuracy of 93.70%.

Pacheco et al. [2] proposed an ensemble of thirteen different convolutional neural networks (CNN) for diagnosing eight skin lesions using ISIC 2019 and achieved an accuracy of 89.7%. The test data represents only 10% of the dataset with an ensemble of the best three models based on the balanced accuracy.

Gessert et al. [3] proposed a deep learning method for skin lesion classification with an ensemble of CNNs including EfficientNets, SENet, and ResNeXt WSL, with a multiple model input resolutions and a search strategy. The proposed ensemble achieved an accuracy of 63.6% in the ISIC 2019 challenge test dataset.

Kassem et al. [4] proposed a deep learning method for classifying eight different skin lesions based on the modified GoogleNet architecture by adding more filters to each layer using ISIC 2019 and a multiclass SVM. They achieved an accuracy of 94.92%.

To et al. [5] proposed a multi-step approach to the problem of classification of skin cancer detection. They divided the task into many smaller sub-tasks easier to treat using several CNNs Inception Resnet, Exception Net, and EfficientNet, and data normalization and a dedicated loss functions. The proposed approach achieved an accuracy of 56.3% on the ISIC 2019 challenge test dataset.

Pollastri et al. [6] proposed a classification model based on the ensembling of two baseline models. Each baseline model contains different CNNs such as ResNet-152, DenseNet-201 and SeResNext-101, SeResNext-50, and SeNet-154.

They employed data augmentation techniques to training the model, and achieved an accuracy of 59.3% on the ISIC 2019 challenge test dataset.

With the impressive growth of deep networks, this paper investigates the ability of commonly pre-trained deep convolutional neural network architectures to diagnose various dermoscopic skin lesions by improving diagnosis with a hierarchical CNN model. Unlike the sequential model used in the most state-of-art approaches, we propose a hierarchical approach to classify eight types of skin lesions using ResNet and DenseNet convolutional neural networks on the ISIC 2019 dataset.

The rest of the paper is structured as follows. Section 2 focuses on the methodology adopted for the classification task and describes the various pre-trained CNN we used. Section 3 describes the experimental results. Last Section provides the conclusion of this study.

2 Methodology

The overall steps of our proposed model are shown in Figure 1. In the first step, we divide the dermoscopic images dataset into two parts: the first represents the training data, and the second represents the testing data. The training phase is composed of five independant CNNs organized into a hierarchical tree, and each one performs individually a decision for a specialized and specific type of skin lesion.

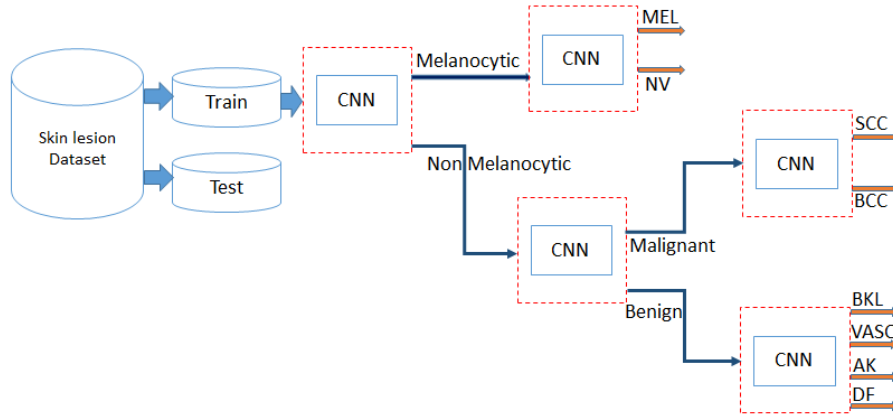


Fig. 1. Proposed method’s flowchart

The first CNN is used for distinguishing between melanocytic and non-melanocytic lesions. Identified Melanocytic lesions are injected into the second CNN to distinguish between melanocytic malignant (Melanoma) and melanocytic benign (Nevus). Contrariwise, identified non-melanocytic lesions are fed to the

third CNN to distinguish between non-melanocytic malignant and non-melanocytic benign lesions. The fourth CNN receives non-melanocytic malignant lesions as input and distinguishes Squamous cell Carcinoma (SCC) from Basal carcinoma (BCC).

Finally, the fifth CNN acquires non-melanocytic benign as input and distinguishes Dermatofibroma (DF), vascular lesion (VASC), actinic keratosis (AK), Benign keratosis (BKL). Once the training phase is finished we evaluate the model with a test dataset.

2.1 Convolutional Neural Networks

In recent years, deep Convolutional Neural Networks (CNN) have enabled many breakthroughs for different applications in the domain of dermoscopy image analysis. It was used as a feature extractor directly from images or for the classification and segmentation tasks [7],[8],[9],[10], with various architectures (LeNet, AlexNet, VGGNet, ResNet, DenseNet, GoogLeNet...etc.). CNN can be trained from scratch, used from pre-trained architectures initially trained on vast datasets, or fine-tuned for the problem under study. In this work, we tackled the skin lesion classification problem by using two of the most accurate pre-trained CNNs among the top-ranked architectures of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [11]: ResNet and DenseNet.

ResNet is a convolutional neural network architecture proposed by He et al. in 2015 [12]. They are characterized by residual blocks that introduce skip connections between layers where each layer feeds into the following layer (Figure 2). ResNet architecture contains 3×3 convolutions, global average pooling, and max-pooling layers, residual blocks, batch normalization layers followed by a fully connected layer, and finally a softmax for the classification.

The ResNet architecture comes with several variations with different numbers of layers such as ResNet18, ResNet34, ResNet50, ResNet101, and ResNet152.

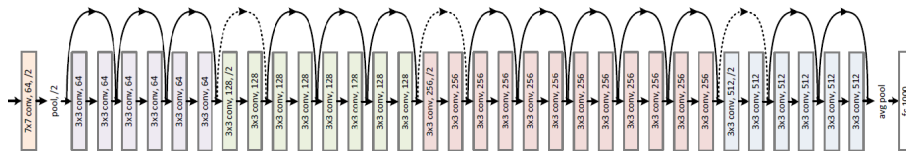


Fig. 2. ResNet architecture [12].

DenseNet is a convolutional neural network architecture proposed by Huang et al. in 2016 [13]. It introduces shortcut connections, where the concept consists of dense blocks linked by transition layers. Each dense block is composed of convolution layers (Figure 3). It is connected to all prior layers, and all feature maps from previous layers are passed to all subsequent layers in the same block. DenseNet architecture contains a convolution filter, global average pooling and

max-pooling layers, transition layers, dense blocks followed by fully connected layers, and finally a softmax for the classification. DenseNet architecture comes with several variations with different numbers of layers such as DensNet-121, DensNet-169, DensNet-201, and DensNet-246.

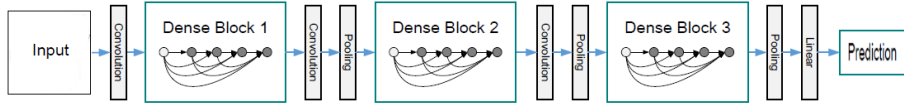


Fig. 3. DenseNet architecture [13].

3 Experimental Results and Discussion

3.1 Dataset

The ISIC 2019 dataset was provided from the International Skin Imaging Collaboration (ISIC) Archive [14], [15], [16] as part of the ISIC 2019 challenge, which maintains an archive of dermoscopic images. This dataset is divided into eight classes: Melanoma (MEL); Melanocytic nevus (NV); Basal cell carcinoma (BCC); Actinic keratosis (AK); Benign keratosis (BKL); Dermatofibroma (DF); Vascular lesion (VASC); Squamous cell carcinoma (SCC). Figure 4 shows some samples from the ISIC 2019 dataset.

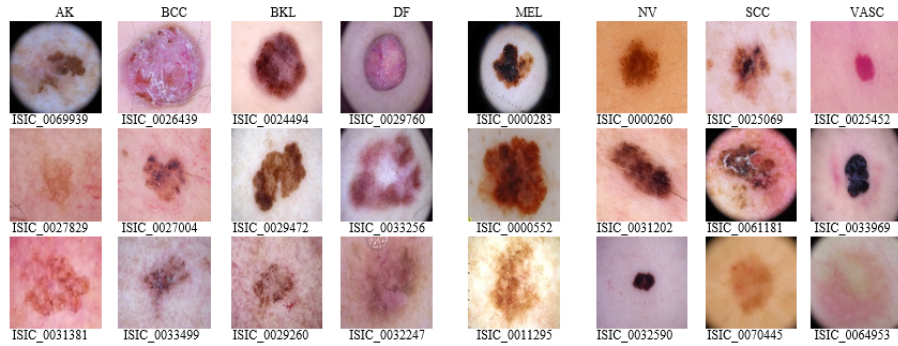


Fig. 4. Samples from ISC 2019 dataset.

3.2 Evaluation Metrics

The evaluation of the proposed model is computed using different performance metrics: accuracy, sensitivity, specificity, and precision, as defined below.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)} \quad (1)$$

$$Sensitivity = \frac{TP}{(TP + FN)} \quad (2)$$

$$Specificity = \frac{TN}{(FP + TN)} \quad (3)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (4)$$

Where TP, FP, FN, TN refer to true positive, false positive, false negative, and true negative respectively.

3.3 Results and Discussion

Several experiments have been conducted with the ISIC 2019 dataset. The dataset was divided into three separated parts: 80%, 10%, and 10%, respectively for the training, validation, and testing datasets. We specify that no data pre-processing has been done. All experiments were trained with the same training dataset and tested with the same test dataset.

All Experiments were performed on a desktop computer equipped with a Core i9 processor, and 32 GB RAM carried on NVIDIA GTX1080ti GPU card. CNNs are trained and evaluated using Python 3.6.10, Keras 2.2.4 [17] and a tensorflow backend 1.14 [18].

We carried out experiments with two scenarios: sequential and hierarchical classification approaches. Each scenario is tested without data augmentation with two CNNs: DenseNet201 and ResNet50. For the sequential approach, we use the CNN directly to perform the classification in end-to-end. In the hierarchical approach, we train five CNNs where each one performs individually a classification for a specific type of skin lesion in a hierarchically structured way such that:

- The first CNN is used for distinguishing between melanocytic and non-melanocytic
- For Melanocytic, the second CNN is used for distinguishing between Melanoma and Nevus
- For non Melanocyt, the third CNN is used to distinguish between non-melanocytic malignant and non-melanocytic benign
- For non Melanocyt malignant, the fourth CNN is used to distinguish between Squamous cell Carcinoma and Basal carcinoma
- For non Melanocyt benign, the fifth CNN is used to distinguish between Dermatofibroma, vascular lesion, actinic keratosis and benign keratosis.

Table 1 depicts the performance results obtained for classifying the ISIC 2019 dataset into eight classes with ResNet50 and DenseNet201 for both sequential and hierarchical architectures without augmentation. Through the obtained results from Table 1, we can observe that the proposed hierarchical model achieves the best results as compared to the sequential model for both architectures DenseNet201 and ResNet50. On the other side, we find that the DenseNet201 model provides the best results as compared to ResNet50.

Table 1. Summary of results obtained with sequential and hierarchical Resnet50 and DenseNet201 architectures.

Metric	Sequential ResNet50	Hierarchical ResNet50	Sequential DenseNet201	Hierarchical DenseNet201
Accuracy	64.03%	77.69%	66.69%	79.43%
Sensitivity	37.33%	76.6%	38.21%	76.37%
Specificity	94.32%	95.83%	94.40%	95.89%
Precision	35.14%	66.35%	35.86%	69.14%

The dataset used in this work is extremely unbalanced. There are many more samples in some classes than in others. For example, As shown in Table 2, NV contains 12875 samples versus DF having only 239 samples. Table 2 shows the distribution of all the samples for each class in the dataset.

Table 2. Distribution of ISIC 2019 samples for each class.

Diagnostic Class	Description	Number of images	Percentage of images
NV	Melanocytic nevus	12875	50.83%
MEL	Melanoma	4522	17.85%
BCC	Basal cell carcinoma	3323	13.14%
SCC	Squamous cell carcinoma	628	2.48%
DF	Dermatofibroma	239	0.94%
VASC	Vascular lesion	253	1.00%
AK	Actinic keratosis	867	3.42%
BKL	Benign keratosis	2624	10.36%

The popular procedure employed to deal with this issue is to artificially augment the less frequent classes by augmenting the data with different ratios for different classes, or to decrease the most frequent classes by upsampling and downsampling or both. In this work, we use both upsampling and downsampling before feeding the images to the networks. The training data was also augmented using the following transformations: rotations, zooming, shearing, Flips (top-

bottom, left-right), and skew-left-right, contrasting. Examples can be seen in Figure 5.

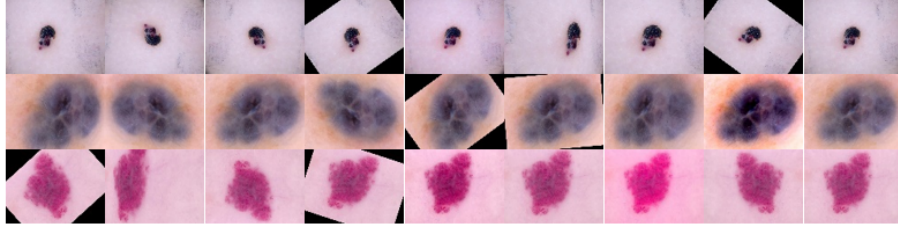


Fig. 5. The different types of data augmentation we have considered.

Tables 3 depicts the results obtained for classifying the ISIC 2019 dataset into eight classes with ResNet50 and DenseNet201 for both sequential and hierarchical architectures with augmentation.

Table 3. Summary of results obtained with sequential and hierarchical ResNet50 and DenseNet201 architectures with data augmentation.

Metric	Sequential ResNet50	Hierarchical ResNet50	Sequential DensNet201	Hierarchical DensNet201
Accuracy	77.81%	78.91%	74.37%	81.11%
Sensitivity	71.52%	67.43%	73.12%	70.59%
Specificity	95,84%	95,88%	95.34%	96.06%
Precision	63.43%	61.29%	60.96%	67.70%

Through the evaluation and results obtained from Tables 3 and 1, we can notice that data augmentation improves the accuracy for all models either sequential or hierarchical. Figures 6 and 7 present a visualization of the performance metrics obtained with DenseNet201 and ResNet50 respectively for sequential and hierarchical approaches.

4 Conclusion

This paper has presented a structured hierarchical classification for multi-class skin lesions based on a deep convolutional neural network system. Further, we presented comparative studies between the proposed hierarchical model and a sequential model using different CNN architectures: Resnet50 and DenseNet201. In this study, the results obtained exhibit that the hierarchically structured classification based on a distinction between melanocytic and non-melanocytic

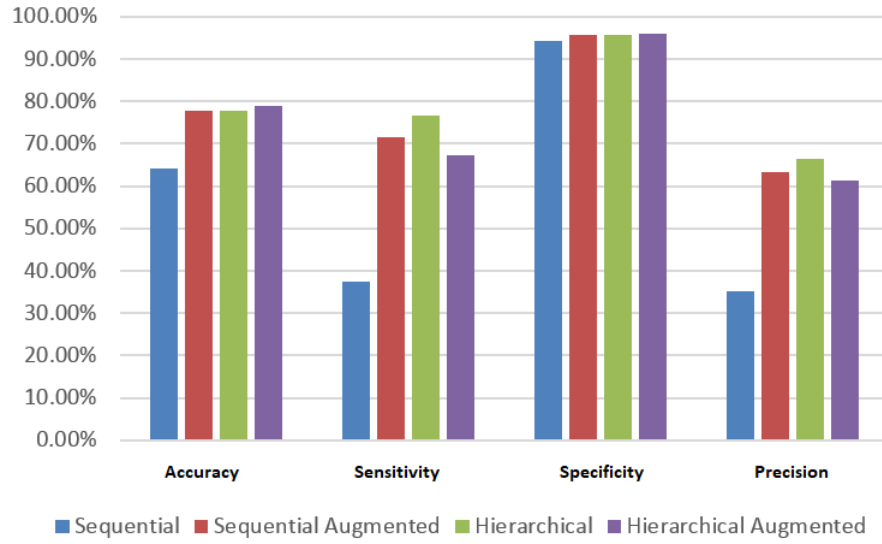


Fig. 6. Summary of results obtained with DenseNet201.

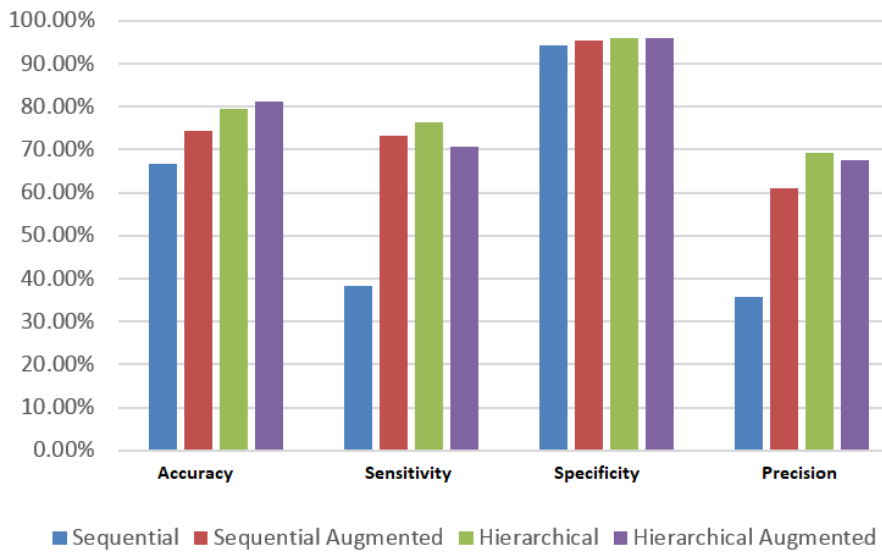


Fig. 7. Summary of results obtained with ResNet50.

followed by a hierarchical diagnosis of the other classes leads to better results than the sequential models.

Acknowledgements This work was completed as part of the Hubert Curien Partnership (PHC) TASSILI cooperation program between Algeria and France under the project code 19MDU212.

References

1. Ahmed, S. A. A., Yanikolu, B., Gksu, O., Aptoula, E.: Skin lesion classification with deep cnn ensembles. In 28th IEEE Signal Processing and Communications Applications Conference (SIU) pp. 1-4 (2020).
2. Pacheco Andre G.C., Ali, A., Trappenberg, T.: Skin cancer detection based on deep learning and entropy to detect outlier samples (2019).
3. Gessert, N., Nielsen, M., Shaikh, M., Werner, R., Schlaefer, A.: Skin lesion classification using ensembles of multi-resolution EfficientNets with metadata. *MethodsX*, 7, 100864 (2020).
4. Kassem, M.A., Khalid, M. H., Mohamed M. F.: Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning. *IEEE Access* 8: pp.114822-114832 (2020).
5. To, T. D., Lan, D. T., Nguyen, T. T. H., Nguyen, T. T. N., Nguyen, H. P., Phuong, L., Nguyen, T. Z.: Multi-Step Skin Cancer Classification. ISIC 2019 Challenge Submission (2019).
6. Pollastri, F., Maronas, J., Parreno, M. Bolelli, F., Paredes, R., Grana, C., Albiol, A.: AImageLab-PRHLT. ISIC Challenge 2019 (2019).
7. Adegun A, Viriri S. Deep learning techniques for skin lesion analysis and melanoma cancer detection: a survey of state-of-the-art. *Artificial Intelligence Review*. 54(2): pp. 811-41 (2021).
8. Benyahia, S. Meftah, B.: Automatic diagnosis system for classification, segmentation, detection and tracking of Skin lesion based on Deep convolutional neural networks. *Models & Optimisation and Mathematical Analysis Journal* Vol.06 Issue 01 (2018).
9. Saket, C.S., Tembhurne, J. V., Diwan, T: A multi-class skin Cancer classification using deep convolutional neural networks. *Multimedia Tools and Applications*. 79.39: pp. 28477-28498 (2020).
10. Khushboo, M., Elahi, H., Ayub,A., Frezza, F., Rizzi, A.: Cancer diagnosis using deep learning: a bibliographic review. *Cancers* 11(9): 1235 (2019).
11. Olga, R., Deng, J. Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z.: Imagenet large scale visual recognition challenge. *International journal of computer vision* 115(3): 211-252 (2015).
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778 (2016)
13. Huang, G., Liu, Z., Weinberger, K. Q.:Densely Connected Convolutional Networks, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2261-2269 (2017)
14. Tschandl P., Rosendahl C., Kittler H.: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data* 5, 180161 (2018)

15. Noel C. F. Codella, David Gutman, M. Emre Celebi, Brian Helba, Michael A. Marchetti, Stephen W. D., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H, Halpern, A: Skin Lesion Analysis Toward Melanoma Detection: A Challenge 2017 International Symposium on Biomedical Imaging (ISBI), International Skin Imaging Collaboration (ISIC) (2017)
16. Combalia, M. Noel C. F. Codella, Rotemberg, V., Helba, B., Vilaplana, V. Reiter, O, Halpern, A.C. Puig, S., Malvey, J.: BCN20000: Dermoscopic Lesions in the Wild (2019)
17. Chollet, F., et al. : Keras. URL: <https://github.com/fchollet/keras> (2015)
18. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M.: Tensorflow: A system for large-scale machine learning. In: 12th fUSENIXg Symposium on Operating Systems Design and Implementation (fOSDIg 16), pp. 265283 (2016)