A Systematic Investigation on Deep Architectures for Automatic Skin Lesions Classification

P. Carcagnì, M. Leo, A. Cuna, G. Celeste, C. Distante (pierluigi.carcagni@cnr.it)

1. Introduction and motivations

- ✓ One of the hot topics in the healthcare and medical fields nowadays concerns the automatic classification of skin lesions.
- ✓ Promising works lever Convolutional Neural Networks (CNN).
- ✓ Existing pipelines mainly rely on complex data pre-processing.
- ✓ There is no systematic investigation about how available deep models can actually reach the accuracy needed for real applications.

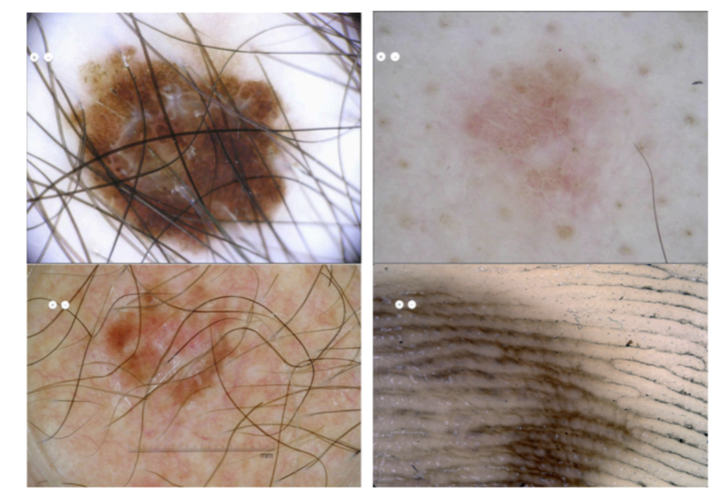
2. Method

- ✓ An end-to-end pipeline has been employed in order to investigate and compare 5 families of CNN architectures: ResNet [4], DenseNet [5], SENet [6] and EfficientNet [7] and RegNet [8].
- ✓ The employed approach performs classifications without the necessity of any preprocessing of the input images (such as for example an image segmentation step), but analyzing only color pixel values.
- ✓ RegNet network design paradigm is for the first time exploited in this application context.

3. Data employed

The dataset used in this work is that provided by the ISIC for the 2019 challenge [1-3], consisting of 25331 images covering 8 types of skin lesions:

- Actinic Keratoses (AKIEC)
- Basal cell carcinoma (BCC)
- Melanocytic nevi (NV)
- Benign keratosis (BKL)
- Vascular skin lesions (VASC)
- Dermatofibroma (DF)
- Melanoma (MEL)
- Squamous cell carcinoma (SSC).



Some challenging situations to be solved while analyzing skin lesions.

MEL	SCC	AK	DF	BKL	VASC	BCC	NV	Total
4522	628	867	239	2624	253	3323	12875	25331

Number of images for each class.

4. Results

Model	F1-score	Precision	Recall
DenseNet121	75.0 ± 1.2	77.6 ± 0.9	72.8 ± 1.7
DenseNet169	78.6 ± 0.8	81.1 ± 0.5	76.6 ± 1.2
DenseNet201	81.7 ± 0.6	**83.6 ± 0.4**	80.1 ± 0.9
EfficientNet-B1	75.0 ± 1.3	76.5 ± 1.0	73.8 ± 1.6
EfficientNet-B2	75.9 ± 1.2	76.4 ± 1.0	75.6 ± 1.3
EfficientNet-B3	79.4 ± 0.8	79.1 ± 0.8	79.8 ± 1.0
EfficientNet-B4	77.7 ± 1.0	78.2 ± 0.9	77.4 ± 1.2
EfficientNet-B5	79.3 ± 0.8	80.0 ± 0.7	78.8 ± 0.9
EfficientNet-B6	79.5 ± 0.9	80.1 ± 0.9	79.0 ± 0.9
EfficientNet-B7	81.4 ± 0.7	81.4 ± 0.6	**81.7 ± 0.7**
RegNetY12GF	81.1 ± 0.7	83.2 ± 0.5	79.2 ± 1.1
RegNetY16GF	**81.9 ± 0.6**	83.4 ± 0.4	80.6 ± 0.8
RegNetY32GF	81.7 ± 0.6	83.5 ± 0.5	80.1 ± 0.8
RegNetY4.0GF	78.6 ± 0.8	80.7 ± 0.6	76.8 ± 1.2
RegNetY6.4GF	81.1 ± 0.6	82.7 ± 0.4	79.7 ± 0.9
RegNetY8.0GF	80.5 ± 0.8	82.0 ± 0.6	79.2 ± 1.0
ResNet101	78.4 ± 0.9	80.0 ± 0.6	77.2 ± 1.1
ResNet152	78.8 ± 0.9	81.0 ± 0.7	77.0 ± 1.1
ResNet50	77.7 ± 1.0	79.9 ± 0.7	75.8 ± 1.3
SE-ResNet101	78.9 ± 0.8	81.1 ± 0.6	77.2 ± 1.1
SE-ResNet50	78.7 ± 1.0	80.7 ± 0.7	77.1 ± 1.3

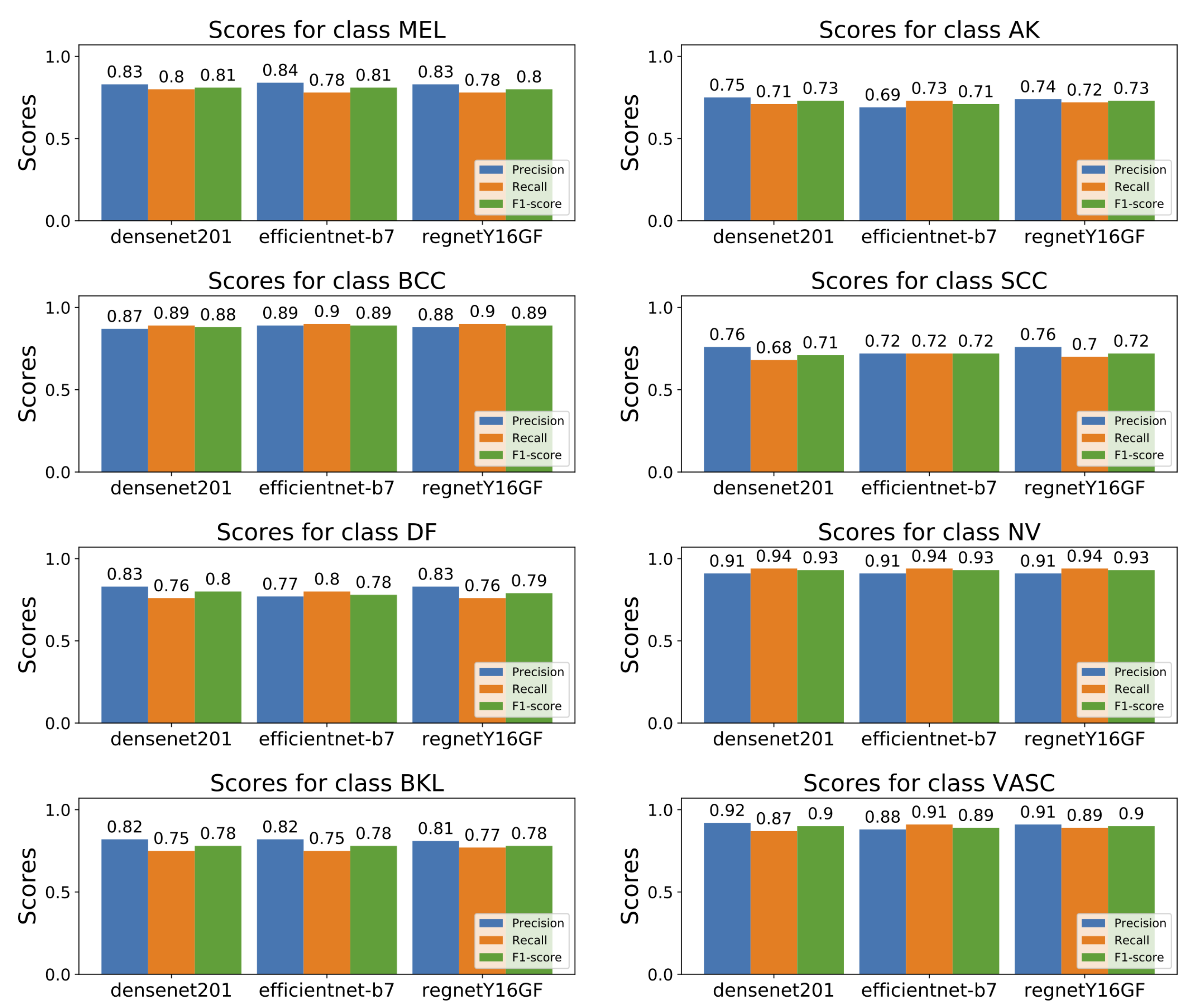
Results (in percentage) gathered on the ISIC 2019 dataset by investigated CNN models. To increase readability the percentage symbol has been omitted.

MEL	DenseNet201, EfficientNet-B7 and RegNetY16GF
AK	RegNetY16GF
BBC	EfficientNet-B7
SCC	RegNetY16GF
DF	DenseNet201
NV	DenseNet201, EfficientNet-B7 and RegNetY16GF
BKL	RegNetY16GF
VASC	RegNetY16GF

The best performing models according to f1-score for the cognition of each ISIC 2019 class.

Model	F1-score	Precision	Recall
Proposed			
DenseNet201	81.7 ± 0.6	**83.6 ± 0.4**	80.1 ± 0.9
EfficientNet-B7	81.4 ± 0.7	81.4 ± 0.6	**81.7 ± 0.7**
RegNetY16GF	**81.9 ± 0.6**	83.4 ± 0.4	80.6 ± 0.8
previous			
Optimal Ensemble [9]			72.5 ± 1.7
Dynamic Augmentation [10]	64		65
Two-level ensembling [11]			59.1

Comparison of the best proposed models with leading approaches in the literature using the same benchmark dataset and the same k-fold strategy.



Densenet-201, EfficientNet-B7 and RegnetY16GF ISIC 2019 classes performance comparison. F1-score (green bars), Precision (blue bars) and Recall (orange bars) are reported for each architecture and each class.

5. Conclusions

Gathered classification results by introduced models were very encouraging and largely outperformed previous approaches on the same dataset. This could be a useful step towards the implementation of an accurate classification system able to supply support in different assistive applications. Future works will explore other deep learning models and ensembling strategies. Conventional ensembling approaches which involve training the base learners separately either in parallel or in a cascading manner will be contemplated.

References

(1) 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 doi.10.1038/sdata.2018.161 (2018)

(2) Noel C. F. Codella, David Gutman, M. Emre Celebi, Brian Helba, Michael A. Marchetti, Stephen W. Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, Allan Halpern: "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)", 2017; arXiv: 1710.05006.

(3) Marc Combalia, Noel C. F. Codella, Veronica Rotemberg, Brian Helba, Veronica Vilaplana, Ofer Reiter, Allan C. Halpern, Susana Puig, Josep Malvehy: "BCN20000: Dermoscopic Lesions in the Wild", 2019; arXiv:1908.02288.

(4) K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

(5) 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.

(6) 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.

(7) M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," arXiv preprint arXiv:1905.11946, 2019.

(8) I. Radosavovic, R. P. Kosaraju, R. Girshick, K. He, and P. Dolla´r, "Designing network design spaces," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 10 428–10 436.

(9) N. Gessert, M. Nielsen, M. Shaikh, R. Werner, and A. Schlaefer, "Skin lesion classification using ensembles of multi-resolution efficientnets with meta data," MethodsX, p. 100864, 2020.

(10) T. A. Putra, S. I. Rufaida, and J.-S. Leu, "Enhanced skin condition prediction through machine learning using dynamic training and testing augmentation," IEEE Access, vol. 8, pp. 40 536–40 546, 2020.

(11) S. Bagchi, A. Banerjee, and D. R. Bathula, "Learning a meta-ensemble technique for skin lesion classification and novel class detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 746–747.