Classification of Skin Lesions by Combining Multilevel Learnings in a DenseNet Architecture

Pierluigi Carcagnì[(✉)], Marco Leo, Andrea Cuna, Pier Luigi Mazzeo,
Paolo Spagnolo, Giuseppe Celeste, and Cosimo Distante

CNR-ISASI, Ecotekne Campus via Monteroni snc, 73100 Lecce, Italy
`pierluigi.carcagni@cnr.it`

Abstract. Automatic recognition and classification of skin diseases is an area of research that is gaining more and more attention. Unfortunately, most relevant works in the state of the art deal with a binary classification between malignant and non-malignant examples and this limits their use in real contexts where the classification of the specific pathology would be very useful. In this paper, a convolutional neural network (CNN) based on DenseNet architecture has been introduced and exploited for the automatic recognition of seven classes (Melanoma, Melanocytic nevus, Basal cell carcinoma, Actinic keratosis, Benign keratosis, Dermatofibroma, Vascular) of epidermal pathologies starting from dermoscopic images. Specialized network architecture and an innovative multilevel fine-tuning method that generates a set of specialized networks able to provide highly discriminative features have been designed. Finally, an SVM model is used for the final classification of the seven skin lesions. The experiments were carried out using an extended version of the HAM10000 dataset: starting from the publicly available images, geometric transformations such as rotations, flipping and affine were carried out in order to obtain a more balanced dataset.

Keywords: Deep Learning · Center loss · Skin lesion classification

1 Introduction

The development of systems based on image analysis for the automatic recognition and classification of skin diseases is an area of research that in recent years is gaining more and more attention. It is indeed a challenging multidisciplinary research area in which the application of modern machine learning techniques, with particular attention to those based on Deep Learning methodologies, have made automatic classification systems increasingly performing and attractive for real uses [13].

Automatic systems for the classification of skin lesions are very desired because, on the one hand, can drive the doctor's attention (allowing the screening of a larger number of patients in the same portion of time) and, on the

other hand, can even allow the development of domestic tools to identify persons most at risk. In the scientific literature, several computer vision based approaches making use of handcrafted features for skin lesions classification can be found [1,4,5,12]. Due to huge inter/intraclass variation and high visual similarity among different classes, the above-mentioned approaches were not able to get satisfying performance in terms of accuracy. In the last years, Deep Convolutional Neural Networks (DCNN) have increasingly being used in the computer vision field for tasks such as image recognition and classification, showing to exceed human performance. This surprising popularity of CNN pushed some researchers to investigate how they can impact on automatic skin lesion classification. Two relevant works exploiting CNN for skin lesion classification are those proposed in [3,6]. Unfortunately, the work in [6] only deals with the binary classification between malignant and non-malignant examples. The work in [3] proposes a unique model to classify multiple classes of skin lesions and to do that a huge amount of data was used in order to handle the large number of parameters in the model. Authors in [14] propose an approach that combines Deep Learning techniques with a low-level segmentation algorithm to distinguish malignant and benign skins lesions. The starting idea of [21] is not far from this, but here authors perform both the segmentation and the classification stages by means of very deep networks with the goal of obtaining more discriminative features for more accurate recognition. The typical degradation problem that occurs when a network goes deeper is overcome by utilizing residual learning technique [7]. In [18] an additional class, representing the visual patterns of regions outside the lesion to reduce their influence on the classification decision, is introduced. In [20] multiple imaging modalities together with patient metadata are provided to a deep neural network to improve the performance of automated diagnosis of five classes of skin cancer. The same five classes are the focus of the approach proposed in [11]. Recently, an interesting comparison between the performance of human experts and Convolutional Neural Networks for skin lesion detection has been proposed in [16].

In this paper, a convolutional neural network (CNN) based on DenseNet architecture [8] has been introduced and exploited for the automatic recognition of seven classes of epidermal pathologies starting from dermoscopic images. In particular, a network architecture more suited to the problem and an innovative multilevel fine-tuning method that generates a set of specialized networks able, also thanks to the linear combination of *soft-max* and *center-loss* [19], to provide highly discriminating features have been designed. To the best of our knowledge, the use of a model ensemble built by starting from one network architecture and generating from that a set of specialized networks through a multilevel fine-tuning method is the main contribution of the paper.

Starting from dermoscopic images, through each network new features are obtained. Features are then concatenated and supplied as input to an SVM model [15] for the final classification of seven skin lesions: Melanoma, Melanocytic nevus, Basal cell carcinoma, Actinic keratosis, Benign keratosis, Dermatofibroma, Vascular. The experiments were carried out using an extended version of

the HAM10000 dataset [17]. In particular, starting from the publicly available images, geometric transformations such as rotations, flipping and affine were carried out in order to obtain a dataset that is as balanced as possible.

The rest of the paper is organized as follow. In Sect. 2 the proposed classification system is described whereas Sect. 3 reports the experimental results. Finally, Sect. 4 concludes the paper and give a glimpse of future works.
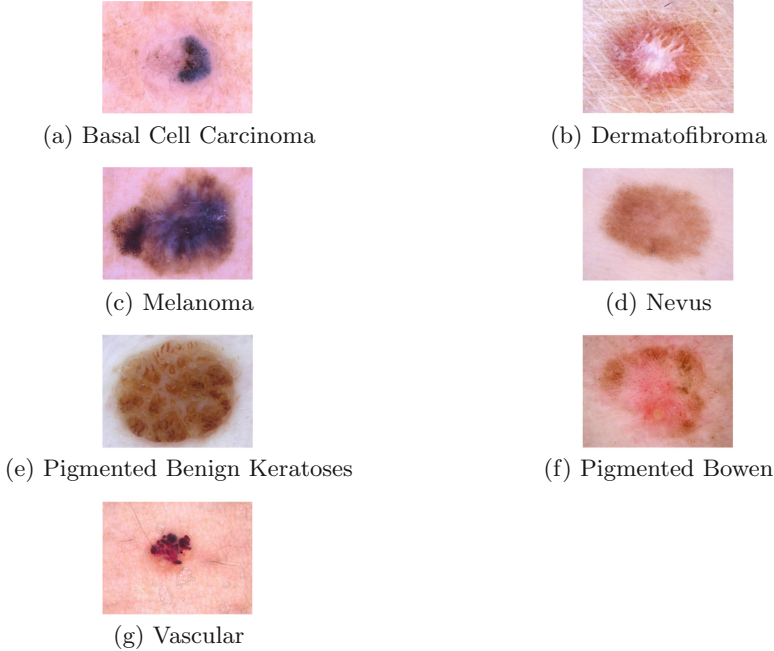
2 Methodology

The challenging problem of the classification of seven classes of skin lesions has been faced by a novel CNN architecture. The architecture was designed by taking as a starting point the one of the Densenet-121 presented in [8]. From the original implementation, the first two Transition Layers and the first two Dense Blocks were maintained whereas the number of layers in the third Dense Block was reduced. The initial internal parameters of the net were set up as provided after the pre-training on the IMAGENET dataset [2]. The third dense block was simplified by reducing the number of its layers and the whole architecture is reported in Table 1.

Table 1. The employed CNN architecture

Layers	Output size	Densenet-62
Convolution	112×112	7×7 conv, stride 2
Pooling	56×56	3×3 max pool, stride 2
Dense Block (1)	56×56	$\begin{bmatrix} 1 \times 1\,\text{conv} \\ 3 \times 3\,\text{conv} \end{bmatrix} \times 6$
Transition Layer (1)	56×56 28×28	1×1 conv 2×2 average pool, stride 2
Dense Block (2)	28×28	$\begin{bmatrix} 1 \times 1\,\text{conv} \\ 3 \times 3\,\text{conv} \end{bmatrix} \times 12$
Transition Layer (2)	28×28 14×14	1×1 conv 2×2 average pool, stride 2
Dense Block (3)	14×14	$\begin{bmatrix} 1 \times 1\,\text{conv} \\ 3 \times 3\,\text{conv} \end{bmatrix} \times 11$
Classification Layer	1×1	7×7 global average pool 7D fully-connected, softmax

The resulting CNN was then fine-tuned on the *HAM10000 Dataset* [17] that consist of $10,015$ dermoscopic images regarding the considered classes of skin lesions: melanoma (MEL), melanocytic nevus (NV), basal cell carcinoma (BCC), actinic keratosis (AKIEC), benign keratosis (BKL), dermatofibroma (DF), vascular (VASC). Some representative samples in the aforementioned dataset are reported in Fig. 1.

(a) Basal Cell Carcinoma

(b) Dermatofibroma

(c) Melanoma

(d) Nevus

(e) Pigmented Benign Keratoses

(f) Pigmented Bowen

(g) Vascular

Fig. 1. One representative example for each of the seven Skin Lesion Classes of the HAM10000 dataset.

Besides, in order to obtain a higher discriminative CNN model, in the learning phase the center-loss function based approach, proposed in [19], was exploited. In particular, in the course of CNN training, high discriminative features are learned considering jointly softmax and center loss functions balanced by means of a hyper parameter. The center loss function was defined by:

$$L_{center} = \frac{1}{2} \sum_{i=1}^{m} \|\mathbf{x}_i - \mathbf{c}_{y_i}\|_2^2 \qquad (1)$$

where the term $\mathbf{c}_i \in \Re^d$ denotes the y_ith class center of deep features $\mathbf{x}_i$.

Finally, the total loss function was defined as linear combination of soft-max L_s and center-loss L_c functions as following:

$$L = L_s + \gamma L_c \qquad (2)$$

where the term γ is a scalar used for balancing the two loss functions. Intra-class minimizations, during the learning phase, were controlled by means of L_c, inter-class maximizations by means of L_s.

3   Experimental Results

The CNN introduced in Sect. 2 was trained by employing a $k-fold$ approach with $k = 5$. The whole HAM10000 (Table 2) was partitioned into five splits: four

Table 2. HAM10000 Training Dataset. Number of images for each class.

MEL	NV	BCC	AKIEC	BKL	DF	VASC	Total
1113	6705	514	327	1099	115	142	10015

of them were used for training and the remaining one for test. The procedure was iterated in order to cover all the combinations of training and test splits. Given the complexity of the network and the number of parameters to be trained during the fine-tuning procedure, the samples of each training/test split were increased by means of geometric transformations. In particular, from each base image in a split additional images were obtained by rotation, flipping and affine transformations (see examples in Fig. 2) in order to obtain training/test sets having the representatives of each class as balanced as possible.

Each image in the balanced splits was then squared by centered cropping of amplitude equal to the shorter side of the starting image. The resulting patch was subsequently resized to a dimension of 224×224 pixels as requested by the input layer of the network.

(a) Base image

(b) Affine

(c) Rotation

(d) Flipping of the rotated image
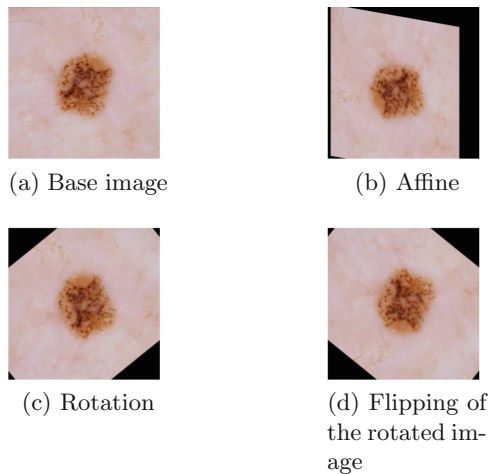
Fig. 2. Image transformations on a Melanoma sample

A multilevel fine-tuning, on the last Dense Block of the employed CNN architecture, was carried out using the training/test data splits provided by the $k - fold$ procedure. In each fine-tuning session, together with the last fully-connected layer, the last two, the last four, the last six, the last eight, the last ten and the last twelve convolution layers of the third Dense Block were modified respectively. To explain better this step, in the first session the last three convolution layers were fine-tuned whereas the parameters in the remaining layers of

the network were not modified. In the second session, the last five convolution layers were fine-tuned and so on. This led to six different models, namely Mod.1, Mod.2, Mod.3, Mod.4, Mod.5, Mod.6 respectively. The classification scores, averaging the obtained $per fold$ results for each model, are in Tables 3 and 4 where Precision/Recall and F1-score are reported.

In Tables 3 and 4 it is evident that each model performs better on a particular class of skin lesions and this experimental evidence led to use an ensemble of nets instead of an end-to-end classifier. To this end, the probability outputs of each network were considered as features and chained in order to obtain a single feature vector in order to represent all the classes as a whole. Feature extraction was then performed on training data and the resulting feature vectors, after dimensional reduction by means of PCA, were used to train a seven class SVM classifier. The obtained SVM model was tested on the test data. The entire procedure was carried out using the $k - fold$ partitioned data and the averaged results related to Precision, Recall, F1-score and confusion matrix of the ensemble classifier are reported in Table 5 and Fig. 3 respectively.

In order to highlight the improvement in the performance of the proposed, the same validation procedure with $k - folded$ splitting of data and the same $center - loss$ approach was carried out by using the Densenet-121 CNN. In this case, the last two convolution layers of the last Dense Block were fine-tuned. Results, in terms of Precision, Recall and F1-Score, of this additional experiment are reported in Table 6. Despite the deeper layout of the net, Densenet-121 CNN showed a worse capacity, in terms of generalization, than the proposed approach to classify the 7 classes of skin lesions.

This can be attributed to the high complexity of the network in terms of the number of layers and to the low numerosity of the dataset used for which the strategy for data augmentation used was not sufficient.

Network training was performed using two NVIDIA GTX 1080Ti cards and the Caffe [9] framework. As optimizer, SGD was chosen with learning rate starting at 0.01, weight decay and momentum equal to 0.0001 and 0.9 respectively. The maximum number of iterations has been set at 75000, decreasing the learning rate by a factor of 10 at each step of 20000 iterations. Finally, the 0.008 value was used for the γ parameter in the Eq. 2. Regarding SVM classifier, an RBF kernel with $\lambda = 0.01$ and $C = 10$ were used.

Experimented outcomes are very encouraging. For all the classes the F1-score was greater than 0.8, except for Melanoma (0.72) and Keratosis (0.62). Since this is a relatively unexplored research field the fair comparison with leading approaches in the literature is not trivial. There is no published work exploiting the HAM10000 dataset indeed. Anyway, it is still possible to get a fair comparison in a quite simple way. In 2018 a dedicated challenge (ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection[1]) was held rightly on the HAM10000 dataset. Task 3 in the challenge was devoted to the seven classes of skin diseases in the HAM10000 dataset. Task 3 was addressed by 141 research groups and, excluding solutions using external data for training, the best one was the ensem-

---

[1] https://challenge2018.isic-archive.com.

Table 3. Results obtained using the 6 fine tuned models (In each row, upper value refers to Precision score, lower value refers to Recall score)

	Mod. 1	Mod. 2	Mod. 3	Mod. 4	Mod. 5	Mod. 6
MEL	0.70	0.70	0.73	0.70	0.75	0.72
	0.56	0.52	0.54	0.60	0.55	0.68
NV	0.91	0.92	0.92	0.94	0.92	0.94
	0.95	0.97	0.97	0.95	0.97	0.96
BCC	0.84	0.83	0.81	0.87	0.83	0.80
	0.82	0.78	0.84	0.80	0.84	0.78
AKIEC	0.79	0.73	0.78	0.76	0.74	0.74
	0.69	0.60	0.56	0.59	0.53	0.53
BKL	0.74	0.76	0.77	0.74	0.78	0.78
	0.72	0.80	0.78	0.82	0.79	0.80
DF	0.80	1.00	0.90	0.75	0.90	0.82
	0.72	0.82	0.82	0.82	0.82	0.82
VASC	0.93	1.0	0.91	0.93	0.92	1.00
	0.93	0.79	0.71	0.93	0.86	1.00
Average	0.81	0.85	0.83	0.93	0.83	0.82
	0.77	0.75	0.75	0.80	0.77	0.80

Table 4. F1-Score results related to the 6 fine tuned models

	Mod. 1	Mod. 2	Mod. 3	Mod. 4	Mod. 5	Mod. 6
MEL	0.62	0.60	0.62	0.65	0.64	0.70
NV	0.93	0.94	0.95	0.94	0.94	0.95
BCC	0.83	0.81	0.83	0.84	0.83	0.80
AKIEC	0.73	0.66	0.65	0.67	0.62	0.62
BKL	0.73	0.78	0.78	0.77	0.79	0.80
DF	0.76	0.90	0.86	0.78	0.86	0.82
VASC	0.93	0.90	0.80	0.93	0.90	1.00
Average	0.79	0.79	0.78	0.80	0.79	0.81

ble of CNN described in [10]. Using PNASNet on 5-fold Validation Data authors in [10] reported a mean precision on 7 classes (namely MCA) of $82,6\% \pm 2.0$ whereas the MCA score, as reported in Table 5 for the ensemble approach proposed in this paper, is 88%.

A final consideration should be made: the approach proposed in this paper is based on an ensemble of models generated by the same reduced network architecture. This leads to reduced models extracted by the same network architecture that can be exploited into embedded systems that are very desirable to quickly move towards portable devices for domestic diagnosis of skin lesions.

Table 5. Precision, Recall and F1-Score results for the proposed ensemble approach.

	Precision	Recall	F1-Score
MEL	0.77	0.68	0.72
NV	0.92	0.98	0.95
BCC	0.91	0.82	0.87
AKIEC	0.80	0.50	0.62
BKL	0.86	0.77	0.81
DF	1.00	0.82	0.90
VASC	0.92	0.86	0.89
Average	0.88	0.76	0.82

Table 6. Precision, Recall and F1-Score results for the original Densenet-121.

	Precision	Recall	F1-Score
MEL	0.60	0.03	0.05
NV	0.95	0.44	0.60
BCC	0.63	0.27	0.38
AKIEC	0.67	0.19	0.29
BKL	0.17	0.93	0.28
DF	0.22	0.18	0.20
VASC	0.21	0.50	0.30
Average	0.49	0.36	0.30
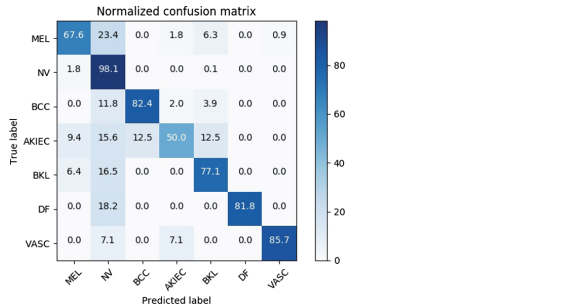
Fig. 3. Ensemble classifier Confusion Matrix

4   Conclusions and Future Work

In this work, a novel approach, based on deep CNNs, for classification of skin lesions has been introduced. It works by using a unique (and not very deep) network architecture from which six models have been generated (each one better performing for specific classes of skin lesions) and then used in an ensemble able

to handle 7 different classes of output. Due to the particular configuration of the CNN, this approach could be exploited into embedded systems that are very desirable to quickly move towards portable devices for domestic diagnosis of skin lesions. Future works will deal with the challenging task of increasing the dataset adding annotated data that can bring to more robust learning of the network parameters. Besides, the possibility to take advantage of some pre-processing step on input images (e.g. colour constancy) will be investigated. Finally, also the use of a preliminary segmentation phase could be considered in order to obtain registered images into a common reference.

References

1. Barata, C., Celebi, M.E., Marques, J.S.: Improving dermoscopy image classification using color constancy. IEEE J. Biomed. Health Inf. **19**(3), 1146–1152 (2015)
2. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 248–255. IEEE (2009)
3. Esteva, A., et al.: Dermatologist-level classification of skin cancer with deep neural networks. Nature **542**(7639), 115 (2017)
4. Garnavi, R., Aldeen, M., Bailey, J.: Computer-aided diagnosis of melanoma using border-and wavelet-based texture analysis. IEEE Trans. Inf Technol. Biomed. **16**(6), 1239–1252 (2012)
5. Glaister, J., Wong, A., Clausi, D.A.: Segmentation of skin lesions from digital images using joint statistical texture distinctiveness. IEEE Trans. Biomed. Eng. **61**(4), 1220–1230 (2014)
6. Haenssle, H., et al.: Man against machine: diagnostic performance of a deep learning convolutional neural network fordermoscopic melanoma recognition in comparison to 58 dermatologists. Ann. Oncol. (2018)
7. 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)
8. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: CVPR, vol. 1, p. 3 (2017)
9. Jia, Y., et al.: Caffe: convolutional architecture for fast feature embedding. arXiv preprint arXiv:1408.5093 (2014)
10. Zhuangy, J., Liy, W.: Skin lesion analysis towards melanoma detection using deep neural network ensemble (2018)
11. Kawahara, J., BenTaieb, A., Hamarneh, G.: Deep features to classify skin lesions. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 1397–1400. IEEE (2016)
12. Kaya, S., et al.: Abrupt skin lesion border cutoff measurement for malignancy detection in dermoscopy images. BMC Bioinf. **17**, 367 (2016)
13. Leo, M., Furnari, A., Medioni, G.G., Trivedi, M., Farinella, G.M.: Deep learning for assistive computer vision. In: Leal-Taixé, L., Roth, S. (eds.) ECCV 2018. LNCS, vol. 11134, pp. 3–14. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11024-6_1

14. Premaladha, J., Ravichandran, K.: Novel approaches for diagnosing melanoma skin lesions through supervised and deep learning algorithms. J. Med. Syst. **40**(4), 96 (2016)
15. Suykens, J.A., Vandewalle, J.: Least squares support vector machine classifiers. Neural Process. Lett. **9**(3), 293–300 (1999)
16. Tschandl, P., et al.: Expert-level diagnosis of nonpigmented skin cancer by combined convolutional neural networks. JAMA Dermatol. **155**(1), 58–65 (2019)
17. Tschandl, P., Rosendahl, C., Kittler, H.: The ham10000 dataset: a large collection of multi-source dermatoscopic images of common pigmented skin lesions. arXiv preprint arXiv:1803.10417 (2018)
18. Vasconcelos, C.N., Vasconcelos, B.N.: Experiments using deep learning fordermoscopy image analysis. Pattern Recogn. Lett. (2017)
19. Wen, Y., Zhang, K., Li, Z., Qiao, Y.: A discriminative feature learning approach for deep face recognition. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9911, pp. 499–515. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46478-7_31
20. Yap, J., Yolland, W., Tschandl, P.: Multimodal skin lesion classification using deep learning. Exp. Dermatol. **27**(11), 1261–1267 (2018)
21. Yu, L., Chen, H., Dou, Q., Qin, J., Heng, P.A.: Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Trans. Med. Imaging **36**(4), 994–1004 (2017)