



Classification of Melanoma Images with Fisher Vectors and Deep Learning

Gastón Liberman^{1,2}, Daniel Acevedo^{1,2}(✉), and Marta Mejail^{1,2}

¹ Facultad de Ciencias Exactas y Naturales, Departamento de Computación,
Universidad de Buenos Aires, Buenos Aires, Argentina
dacevedo@dc.uba.ar

² Instituto de Investigación en Ciencias de la Computación (ICC),
CONICET-Universidad de Buenos Aires, Buenos Aires, Argentina

Abstract. The present work corresponds to the application of techniques of data mining and deep training of neural networks (deep learning) with the objective of classifying images of moles in ‘Melanomas’ or ‘No Melanomas’. For this purpose an ensemble of three classifiers will be created. The first corresponds to a convolutional network VGG-16, the other two correspond to two hybrid models. Each hybrid model is composed of a VGG-16 input network and a Support Vector Machine (SVM) as a classifier. These models will be trained with Fisher Vectors (FVs) calculated with the descriptors that are the output of the convolutional network aforementioned. The difference between these two last classifiers lies in the fact that one has segmented images as input of the VGG-16 network, while the other uses non-segmented images. Segmentation is done by means of an U-NET network. Finally, we will analyze the performance of the hybrid models: the VGG-16 network and the ensemble that incorporates the three classifiers.

Keywords: Melanoma classification · Deep learning · Fisher vectors

1 Introduction

Melanoma corresponds to skin cancer that causes 75% of the deaths of all cutaneous cancer diseases. It is estimated that there are 160000 new cases of melanomas and 48000 deaths per year according to the World Health Organization. Dermatologists perform a biopsy in order to confirm the presence or not of a melanoma. This procedure increases its complexity when the person has a large number of suspicious moles, since it turns out to be an invasive process. In addition, the detection of melanoma is totally conditioned to the training of the dermatologist. Therefore it is important to have a tool that allows to classify the presence of a melanoma without invading the human body and whose effectiveness of classification is superior to the average of the dermatologists, reducing the subjectivity of human vision.

During the last years the use of automatic processes in image classification has increased notably and the classification of melanoma is not the exception.

© Springer Nature Switzerland AG 2019

R. Vera-Rodriguez et al. (Eds.): CIARP 2018, LNCS 11401, pp. 732–739, 2019.

https://doi.org/10.1007/978-3-030-13469-3_85

One of the first works corresponds to the use of the ABCD rule [1]; it is a formula computed from a combination of different characteristics of the mole: asymmetry, edges, color and diameter of the lunar. Depending on the score that is obtained, the injury is classified. Subsequently, the method of Menzies [2] has a set of characteristics related to a benign mole and another that are strictly related to a melanoma. Therefore, depending on the characteristics found in the lesion, it will be considered a malignant or benign mole. Another technique considers seven points (ELM7) [3] in which seven fundamental characteristics of the moles are evaluated. Regarding the segmentation of the lesion the Otsu method was mainly used [4]. Given these three methods and the Otsu method, different automatic melanoma detection systems were developed as can be seen in Table 1. The neural networks were used at first, and then migrated to other data mining techniques. However, from 2012/2013, with the increase in hardware speed, neural networks begin to be used again but this time with more layers. Then deep convolutional networks emerge, with a very good performance in the classification of images; today they determine the state-of-the-art in this field. In addition, papers were published independently in terms of classification and segmentation. Table 2 lists the most important works in this field. The paper is organized as follows. Section 2 describes the methodology. Section 3 presents the experimental setup and results, and finally concluding remarks are given in Sect. 4.

Table 1. Previous background to the use of deep learning methods

References	Method	Segmentation	Classification
H Iyatomi et al. 2005. [5]	ABCD roule	Growing regions	Neural Networks
G. Capdehourat et al. 2009. [6]	ABCD roule and Argenziano method	Otsu method	Decision trees
J. F. et al. 2009. [7]	ABCD roule	Otsu method	Fusion of decision trees and bayesian networks
G. Leo et al. 2010. [8]	Argenziano method	Otsu method	Decision trees
D. Ruiz et al.2011. [9]	ABCD roule	Otsu method	Multilayer perceptron, bayesian classifier, KNN
G. Capdehourat et al. 2011 [10]	Clinic support	Otsu method	Adaboost of decision trees

2 Models Description

In the present work, images from the Buenos Aires Italian Hospital¹, ISIC Archive² and Dermnet³ were used. The automatic classification of melanoma

¹ We thank Victoria Kowalckzuc MD from the Buenos Aires Italian Hospital for her medical advice and providing the mole images.

² <https://isic-archive.com/>.

³ <http://www.dermnet.com/images/Malignant-Melanoma>.

Table 2. Reference works using deep learning methods

References	Type	Method
Xie, et al. 2016. [11]	Classification	SGNN + GA
O. Ronneberger, P. Fischer, and T. Brox et al. 2015. [12]	Segmentation	Convolution and UpSampling (UNET)
N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern, and J. R. Smith, et al. 2015. [13]	Classification	CNN, Sparse Coding and SVM as ensembler
Demyanov, et al. 2016. [14]	Classification	CNN 8 layers
J. Kawahara, et al. 2016. [15]	Classification	CNN converting fully connected layers into convolutional layers to extract features
A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun et al. 2017. [16]	Classification	CNN - Inception V3
Zhen Yu et al. 2017. [17]	Classification	Fisher Vectors, CNN and SVM as classifier

using convolutional networks (CNNs) was investigated by Lecun et al. [18]. Even though they have a very good performance in the classification of the images, they have the disadvantage of being sensitive to the invariance of transformations: rotations, change of scale and orientation. One of the ways to solve this problem is to combine CNN (local descriptors) with Fisher vectors [19]. In this work we propose an ensemble of the hybrid system introduced by Yu et al. [17] with that of a CNN network to improve the performance of the resulting classifier. Our model includes an image preprocessing, a segmentation, and classification modules. In the preprocessing module the images were rescaled and Max-Constancy was applied [20]. Max-Constancy is a technique used to filter the effects of the light source that can produce a distortion in the image (similar to the filter that automatically performs the human being). The segmentation was performed by the UNET network [12]. In order to train the network that segments the lesion, only the images that come from the ISIC Archive source are used since they have the respective masks. The classification was solved by means of the VGG-16 network [21] trained on Imagenet⁴. In addition, heat maps were visualized on the classified images to evaluate the performance in the training phase of the VGG-16 network. The heat maps correspond to the areas that CNN uses to perform the aforementioned classification. The method that will be used to obtain these heat maps is GRAD-CAM [22].

For the CNN model, once the dataset was created data augmentation was carried out to increase the number of images in the training set. That is, they

⁴ <http://www.image-net.org>.

were rotated at 0, 90 and 180°, and zoomed at the same time, producing different versions of each image.

Fine-Tuning was performed for both the CNN model and the hybrid model. That is, the Fully-Connected layers were first removed and the images were passed through the convolutional layers. The output of the network will be called “Deep Features”. Then with these descriptors a mini network formed by the last block of the CNN network and the Fully-Connected layers was trained. On the other hand, for the hybrid model, the descriptors of CNN were encoded in the Fisher vectors, which are the inputs of a SVM with the aim of training it and generating the hybrid classifier. The final classifier is an ensemble of the two hybrid models and a CNN model (VGG-16): the final class probability is estimated by the average of the probabilities provided by each model.

The **Classifier 1** is the simplest model. It is composed of a VGG-16 network pre-trained in IMAGENET, in which the weights of the last block of the network and of the last fully connected layer are changed based on the use of training and validation datasets. The Hybrid Model [17] (**Classifier 2**) takes as input the deep features which are reduced in dimensionality by Principal Component Analysis (PCA). A Gaussian mixture model (GMM) is first learned by sampled images from training set, and the Fisher Vector representation is calculated for each image. The number of PCA and GMM components are learned with the objective of maximizing the ROC curve in the Validation dataset. It should be noted that the input of this classifier corresponds to the descriptors that are the product of passing the images to the last convolutional layer of Classifier 1. This model is shown in Fig. 1. Then we have **Classifier 3** (see top chart in Fig. 3) that takes as input the descriptors of the segmented images for which we train a U-NET network with the objective of segmenting the images. Figure 2 shows an example of the application of the segmenter to an image; the grayscale area is not considered as part of the lesion for the segmenter. It should be noted that **Classifier 3** is similar to **Classifier 2** but the descriptors are obtained from the segmented images instead of the original images. Our proposed model, built by the **Ensemble** of the three classifiers, can be seen at the bottom of Fig. 3.

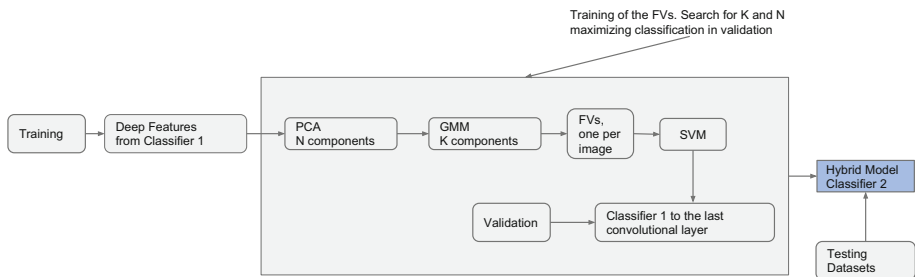


Fig. 1. Scheme of hybrid models

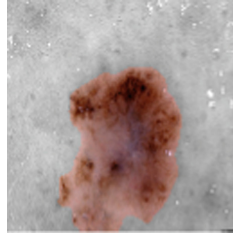


Fig. 2. Segmented melanoma with a gray scale outside the lesion

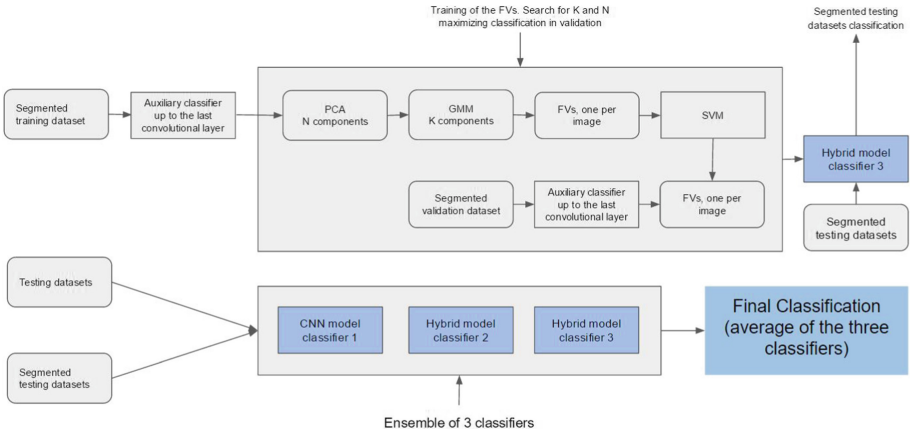


Fig. 3. Hybrid model and the ensemble of three classifiers

Finally, the **Ensemble** of classifiers is applied to the testing dataset (constituted by the original images as segmented). The segmentation in the testing file is necessary to obtain the input of **Classifier 3**, while the input of classifiers 1 and 2 are the original images.

3 Results

Several testing datasets were generated and in each of them the proposed model was applied. The original testing dataset is composed of 100 images where there are 50 melanomas and 50 benign moles, we will call it ISIC testing dataset. Then there is the XY testing dataset, in which operations of translations are carried out on the X and Y axes, obtaining 250 melanomas and 250 benign moles. In addition, there is the ROT testing dataset, consisting of rotation operations and reflection of the original images, obtaining 250 melanomas and 250 benign moles. Finally, we have the testing dataset of the Italian Hospital (HI) with lower quality images in relation to those used in the training and the ISIC, ROT and XY test dataset. As mentioned in the Sect. 2, Max-Constancy was applied to

all the images. Experiments were executed on a machine with core-i7 6700 HQ microprocessor, 16 GB of memory and with a GTX 950m GPU.

The parameters that achieve an optimal performance in each model are shown in Table 3.

Table 3. Training parameters for each model.

Model	Deep learning								FV		SVM(RBF)	
	Optim. U-NET	Batch Size U-NET	Epochs U-NET	Learning Rate U-NET	Optim. VGG-16	Batch Size VGG-16	Epochs VGG-16	Learning Rate VGG-16	PCA (N)	GMM (K)	Cost (C)	Gamma
Classifier 1 (VGG-16)					SGD	16	100	10^{-5}				
Classifier 2 (Hybrid)					SGD	16	100	10^{-5}	40	20	100	$\frac{1}{\#feat.}$
Classifier 3 (Hybrid)	Adam	4	220	10^{-5}	SGD	16	100	10^{-5}	40	6	100	0.1

It should be noted that in the training of the VGG-16 network the learning rate is reduced (to half its value) if the same ROC value is obtained during 5 epochs. The binary cross-entropy is used as loss function while in the U-NET network, the Jaccard coefficient was used to measure the error between the predicted mask and the real mask.

Although with the Hybrid model the lowest performance is obtained, it is observed that when combined with Classifier 1, the best performance is obtained. The latter is visualized in Fig. 4 (a) in relation to (b) and (c) that represent Classifier 1 and the Hybrid model without segmentation (Classifier 2), respectively.

4 Conclusions

In this work we develop techniques of data mining and deep training of neural networks (deep learning) with the objective of classifying images of moles in ‘Melanomas’ or ‘No Melanomas’. An ensemble of three classifiers is utilized for this purpose. The classifiers include a convolutional neural network (VGG-16), Fisher vector encoding, and image segmentation by means of a U-NET network. It was found that when the testing dataset is composed of images that are originated through translation operations on the X and Y axes of the original images, the performance in the CNN network decreases more rapidly than with the hybrid model. This is because the data augmentation performed in the training dataset, used in the training of the network, does not include the translation operation on the axes. We conclude that the CNN network is less invariant than the hybrid model for this type of operation, that was not applied in the data augmentation in the training dataset.

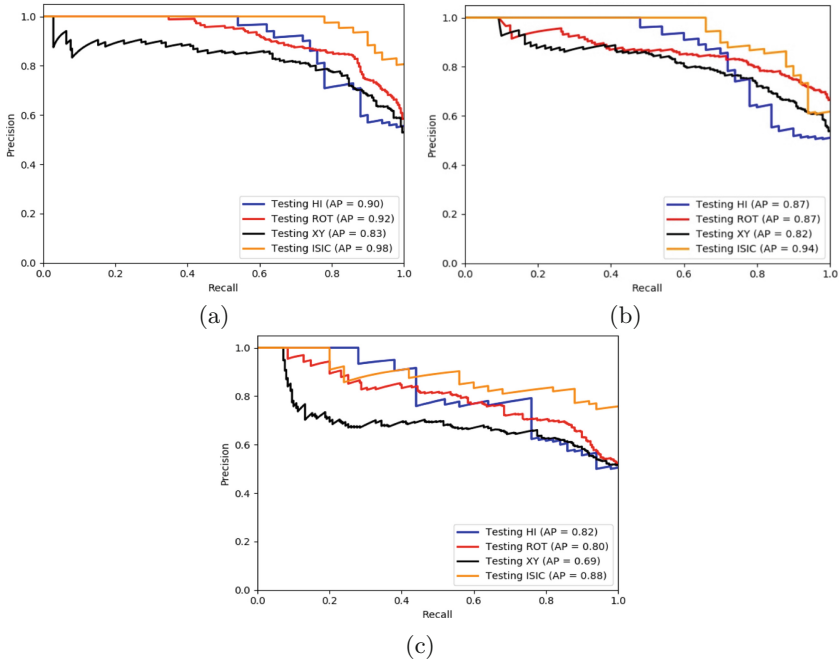


Fig. 4. (a): PR curves in the testing datasets (Ensemble model), (b): PR curves in testing datasets (Classifier 1) and (c): PR curves in testing datasets (Classifier 2).

References

1. Stolz, W., et al.: ABCD rule of dermatoscopy: a new practical method for early recognition of malignant melanoma. *Eur. J. Dermatol.* **4**, 521–527 (1994)
2. Menzies, S.W., Ingvar, C., Crotty, K.A., McCarthy, W.H.: Frequency and morphologic characteristics of invasive melanoma lacking specific surface microscopic features. *Arch. Dermatol.* **132**, 1178–1182 (1996)
3. Argenziano, G., Fabbrocini, G., Carli, P., De Giorgi, V., Sammarco, E., Delfino, M.: Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. *Arch Dermatol.* **134**, 1563–1570 (1998)
4. Otsu, N.: A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **9**(1), 62–66 (1979)
5. Iyatomi, H., Oka, H., Hashimoto, M., Tanaka, M., Ogawa, K.: An internet based melanoma diagnostic system Toward the practical application. In: *Proceedings of the 2005 IEEE Symposium on Computational Intelligence in Bioinformatics and Comp. Biology*, pp. 1–4 (2005)
6. Capdehourat, G., Corez, A., Bazzano, A., Musé, P.: Pigmented skin lesions classification using dermatoscopic images. In: Bayro-Corrochano, E., Eklundh, J.-O. (eds.) *CIARP 2009. LNCS*, vol. 5856, pp. 537–544. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-10268-4_63

7. Alcon, J.F., et al.: Automatic imaging system with decision support for inspection of pigmented skin lesions and melanoma diagnosis. *IEEE J. Sel. Top. Signal Process.* **3**, 14–25 (2009)
8. Leo, G.D., Paolillo, A., Sommella, P., Fabbrocini, G.: Automatic diagnosis of melanoma: a software system based on the 7-point check-list. In: *HICSS (2010)*
9. Ruiz, D., et al.: A decision support system for the diagnosis of melanoma: a comparative approach. *Expert. Syst. Appl.* **38**, 15217–15223 (2011)
10. Capdehourat, G., et al.: Toward a combined tool to assist dermatologists in melanoma detection from dermoscopic images of pigmented skin lesions. *Pattern Recognit. Lett.* **32**, 1–10 (2011)
11. Xie, F., Fan, H., Yang, L., Jiang, Z., Meng, R., Bovik, A.: Melanoma classification on dermoscopy images using a neural network ensemble model. *IEEE Trans. Med. Imaging* **36**, 849–858 (2016)
12. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation [arXiv:1505.04597](https://arxiv.org/abs/1505.04597) (2015)
13. Codella, N., Cai, J., Abedini, M., Garnavi, R., Halpern, A., Smith, J.R.: Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images. In: Zhou, L., Wang, L., Wang, Q., Shi, Y. (eds.) *MLMI 2015. LNCS*, vol. 9352, pp. 118–126. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24888-2_15
14. Demyanov, S., et al.: Classification of dermoscopy patterns using deep convolutional neural networks. In: *IEEE 13th ISBI (2016)*
15. Kawahara, J., et al.: Deep features to classify skin lesions. In: *ISBI*, pp. 1397–1400 (2016)
16. Esteva, A., et al.: Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**(7639), 115 (2017)
17. Yu, Z., et al.: Hybrid dermoscopy image classification framework based on deep CNN and Fisher vector. In: *14th IEEE ISBI, Melbourne, Australia (2017)*
18. Lecun, Y., et al.: Gradient-based learning applied to document recognition. *Proc. IEEE* **86**(11), 2278–2324 (1998)
19. Sanchez, J., et al.: Image classification with the fisher vector: theory and practice. *IJCV* **105**, 222–245 (2013)
20. Gijssenij, A., et al.: Computational color constancy: survey and experiments. *IEEE Trans. Image Process.* **20**, 2475–2489 (2011)
21. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. [arXiv 1409.1556](https://arxiv.org/abs/1409.1556) (2014)
22. Selvaraju, R., et al.: Grad-CAM: why did you say that? Visual explanations from deep networks via gradient-based localization (2016)