Pixel Classification Methods to Detect Skin Lesions on Dermoscopic Medical Images

Fabrizio Balducci(✉) and Costantino Grana

Dipartimento di Ingegneria "Enzo Ferrari",
Università degli Studi di Modena e Reggio Emilia,
Via Vivarelli 10, 41125 Modena, MO, Italy
{fabrizio.balducci,costantino.grana}@unimore.it

Abstract. In recent years the interest of biomedical and computer vision communities in acquisition and analysis of epidermal images increased because melanoma is one of the deadliest form of skin cancer and its early identification could save lives reducing unnecessary medical treatments. User-friendly automatic tools can be very useful for physicians and dermatologists in fact high-resolution images and their annotated data, combined with analysis pipelines and machine learning techniques, represent the base to develop intelligent and proactive diagnostic systems. In this work we present two skin lesion detection pipelines on dermoscopic medical images, by exploiting standard techniques combined with workarounds that improve results; moreover to highlight the performance we consider a set of metrics combined with pixel labeling and classification. A preliminary but functional evaluation phase has been conducted with a sub-set of hard-to-treat images, in order to check which proposed detection pipeline reaches the best results.

Keywords: Dermoscopic images · Image analysis · Annotation · Usability · Clustering

1 Introduction

The malignant melanoma is one of the most common and dangerous skin cancer, in fact 100,000 new cases with over 9,000 deaths are diagnosed every year by only considering the USA [7,8]; in this context automated system for fast and accurate melanoma detection are well accepted, also considering technical approaches like machine learning methods and the Convolutional Neural Networks.

The premise for this work was the development of an annotation tool for epidermal images: this has been the first step to create an heterogeneous data integrated system which architecture is depicted in Fig. 1.

This architecture presents the Digital Library with inner modules dedicated to image analysis and feature extraction and external tools to manage the *Annotation*, the *Information Visualization* and *Query* search capacities [3], for example by exploiting the DICOM (Digital Imaging and Communications in Medicine) metadata standard.

S. Battiato et al. (Eds.): ICIAP 2017, Part II, LNCS 10485, pp. 444–455, 2017.
https://doi.org/10.1007/978-3-319-68548-9_41

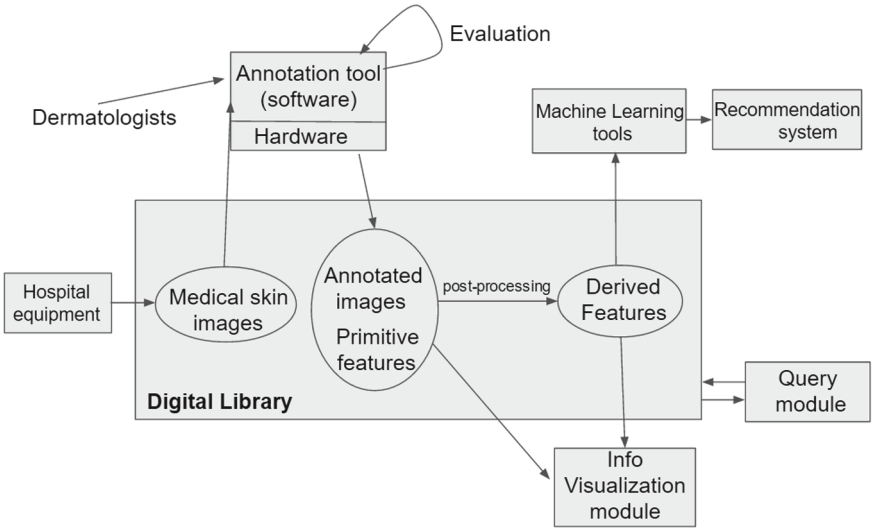

Fig. 1. The system architecture proposed for a medical digital library

The dataset coming from hospital equipment currently consists of 436 dermoscopic skin images in JPEG format with 4000×2664 or 4000×3000 resolution. In the view of the integrated system architecture, at this stage we focus on the exploit of state-of-art reliable methods for extracting the visual features we need, also introducing workarounds to improve results. The annotation software in Fig. 2 has been developed for domain experts like dermatologists which have peculiar working protocols and do not have much time to skill themselves on externals tools since usually overworked.

As hardware platform we chose the *Microsoft Surface Pro 3* which is a powerful non-invasive device that can be used in mobility in medical environments; it ensures that the recognized strokes will be only those that come from the specific *Surface Pen*, avoiding unwanted strokes coming from touch gestures or oversight movements.

This paper is organized as follows: Sect. 2 introduces some literature about epidermal and melanoma image analysis while Sect. 3 explains the skin lesion detection pipelines; Sect. 4 reports results obtained by the experimental sessions and finally, conclusions and future work are illustrated in Sect. 5.

2 Related Work

Nowadays standard video devices are commonly used for skin lesion inspection systems, in particular in the telemedicine field [20]; however, these solutions bring some issues, like poor camera resolution (melanoma or other skin details can be very small), variable illumination conditions and distortions caused by the camera lenses [14]. A complete and rich survey about skin lesion characterization is

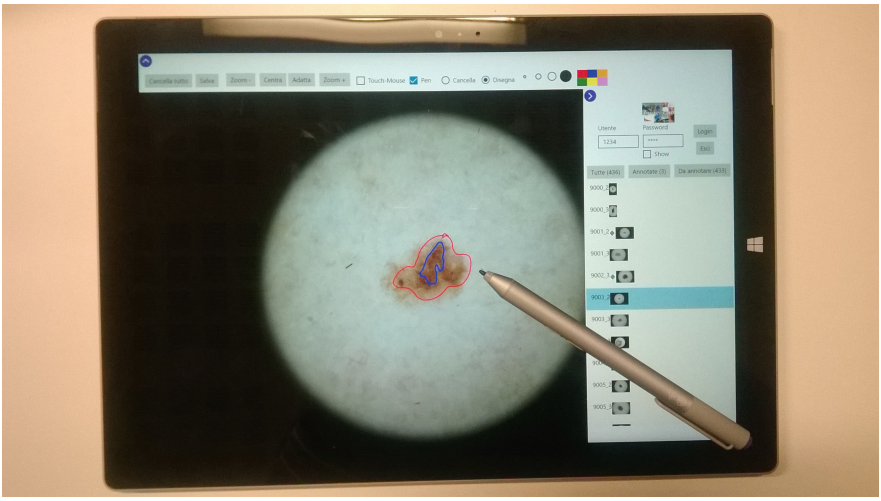

Fig. 2. Microsoft Surface is the hardware platform for the annotation tool (Color figure online)

in [5] with artifact removal techniques, evaluation metrics, lesion detection and preprocessing methods. The work of Seidenari *et al.* [19] provides an overview about the detection of melanoma by image analysis, Wighton *et al.* [21] presents a general model using supervised learning and MAP estimation that is capable of performing many common tasks in automated skin lesion diagnosis and Emre Celebi *et al.* [9] treats lesion border detection with thresholding methods as Fan *et al.* [10] that use saliency combined with Otsu thresholding. The work of Peruch *et al.* [17] faces lesion segmentation through Mimicking Expert Dermatologists' Segmentations (MEDS) and in Liu *et al.* [13] propose an unsupervised sub-segmentation technique for skin lesions.

In Codella *et al.* [6] manually pre-segmented images, already cropped around the region of interest, have been used in conjunction with hand-coded and unsupervised features to achieve state-of-the-art results in melanoma recognition task, with a dataset of 2,000 images. Learning approaches are exploited in Schaefer *et al.* [18] and deep learning techniques in Abbes and Sellami [2]; studies in [12,16] have been exploited in literature, while a combination of hand-coded features, sparse-coding methods, HOG and SVMs are used in Bakheet [4].

Finally, in 2016 a new challenge, called *Skin Lesion Analysis toward Melanoma Detection* [11], has been presented: the aim is to use one of the most complete dataset of melanoma images, collected by the *International Skin Imaging Collaboration* (ISIC) and obtained with the aggregation dataset of dermoscopic images from multiple institutions, to test and evaluate the automated techniques for the diagnosis of melanomas. The ISIC database has been also exploited in Yuan et al. [22] with a 19-layer deep convolutional neural network while classification and segmentation are achieved using deep learning approaches are also in Majtner et al. [15].

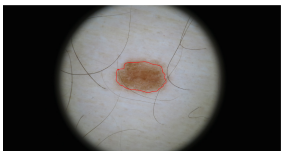

Fig. 3. A dermoscopic image annotated by a dermatologist (Color figure online)

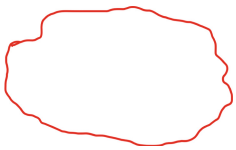

Fig. 4. The extracted annotation (primitive feature) (Color figure online)

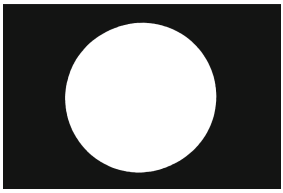

Fig. 5. External black frame pixel mask (first version)

3 Visual Features and Skin Image Processing

In Fig. 3 is depicted a dermoscopic image where a dermatologist annotated the principal external contour of a skin lesion (red color): the primitive meta-data directly extracted by the annotation tool is in Fig. 4.

One of the targets of this system is to create and manage the image dataset for the automated analysis by combining low-level visual features representation, image processing techniques and machine learning algorithms: we want to obtain pipelines that detects skin lesions and automatically draws one or more contours by mimiking and increasing the dermatologist knowledge.

The drawn strokes introduce *primitive features* consisting of binary masks, coordinate points, color codes and pen sizes; the image processing functions extract *derived features* like contours, shapes, intersections, color features and numerical values.

Our method will exploit the manual annotations alongside image processing algorithms with the aim to evaluate how many pixels of the skin lesion could be automatically detected by the system; a pre-processing phase is necessary to remove thick skin hairs, because these artifacts influence shape and contours extraction [23].

3.1 Hair Removing

To accomplish the task we were inspired to the *DullRazor* [1] pipeline consisting of

1. *Detection* step that locates into an RGB image the slender and elongated structures that resemble the hairs on the skin by making an hair pixel mask
2. *Replacement* step that replaces each detected hair pixel with the interpolation of two lateral pixels chosen from a line segment built on a straight direction.

The *Detection* phase exploits the *generalized grayscale morphological closing* operation G_c: for each color channel c (Red-Green-Blue) the operator makes a set of morphologic closing by using different kernels with the aim to compare (by

choosing the highest value c_p), for each pixel, which kernel better approximates a potential hair shape.

The value of G_c for each pixel p is calculated as:

$$\forall c \in r, g, b, \forall p, G_c = |b_c(p) - max(c_p)| \tag{1}$$

where $b_c(p)$ is the actual image pixel value and c_p measures how many pixels of the kernel are verified as hair structure for an image pixel p.

The kernel represents a sort of 'skeleton' that reconstructs pixel-to-pixel the hair shape (elongated, slight, weakly curvilinear) on the kernel-closed image and so the kernel structure is a very critical variable; we used four kernels for each of the possible directions (11 pixels horizontal at 0°/180°, 11 pixels vertical at 90°/270°, 9 pixels oblique at 45°/225° and 9 pixels oblique at 135°/315°) from which a potential hair could spread from a single pixel (located at the center of the kernel).

The final hair mask M is the union of the resulting pixel masks M_r, M_g, M_b, where each mask is obtained by a threshold on the generalized closing operator value previously calculated for each pixel.

Before the *Replacement* phase is performed, we must verify that each candidate hair pixel belongs to a valid thick and long structure so, for each direction previously described, a path is built having the pixel at the center until the non-hair regions are reached. The longest path is used to take the two interpolation pixels, selected on both the sides, perpendicular to the directions, at a fixed distance from the hair structure borders.

This pipeline is greatly affected by the variables used at each step, for example the kernel shape, the G_c threshold for the hair masks and the distance for the choose of the interpolation pixels.

3.2 Skin Lesion Detection

After the treatment that removes most of the hairs, it is possible to design the detection of the skin lesion area: we begin by using standard image processing techniques but with a fine-tuned *work pipelines* that will filter and refine results:

1. Otsu thresholding with pixel mask
2. Color clustering with pixel mask and cluster tolerance.

By considering the peculiar structure of a dermoscopic image, a binary mask M0 (Fig. 5) must be constructed to approximate the large black frame which surrounds the bright circular skin that contains the lesion. We execute a morphologic *erosion* on the original blurred image with a kernel size of 201 × 201: now the Otsu threshold (Fig. 8) is calculated considering the pixels not in M0 and, for future tests and comparisons, the resulting pixels will be in red color (in Fig. 6 the original image, in Fig. 7 with hairs removed).

For the color clustering technique we must develop an heuristic to differentiate if a cluster belongs to the bright skin or to the lesion skin by assign the label *skin* or *lesion* to each cluster; moreover it must be considered that a cluster can

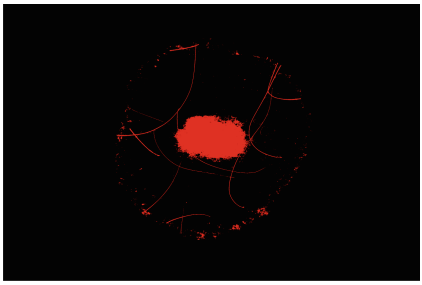

Fig. 6. Otsu with pixel mask, no hairs removed (Color figure online)

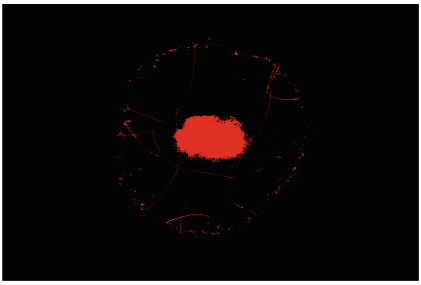

Fig. 7. Otsu with pixel mask, hairs removed (Color figure online)

easily intersect the two area types (some pixels on the lesion and others on the skin) especially on the borders.

To deal with this ambiguity we will develop a pixel *toleration area*: we execute a further morphologic *closing* and *erosion* on the otsu image by using two different kernels to obtain an *enlarged mask* (Fig. 9) that builds safety areas around the image borders (a side effect is that scattered hair pixels will be removed but groups may be enlarged).

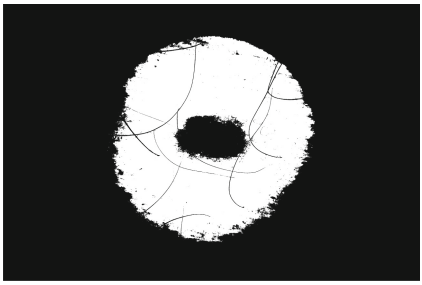

Fig. 8. Otsu from black binary mask, no hairs removed

Fig. 9. Closed and eroded Otsu for the color cluster toleration area

The color cluster will be computed with OpenCV and, as for Otsu pipeline, by considering only the pixels not in the M0 mask; the number of clusters ($K = 10$) has been chosen empirically after a set of experimental sessions. Moreover the enlarged pixel mask in Fig. 9 has been divided into two sub-masks M1 and M2: the M1 mask represents the largest connected component that from now will replace M0, while M2 (the second largest connected component) will be considered as the tolerance area that approximates the central skin lesion pattern and that must contains the bigger and uniform clusters.

For each color cluster we count the number of pixels considered 'skin' or 'lesion' by using the M2 masks so if a cluster has more than 10% of its pixels

out from this tolerance area it will be considered as simple skin (all of its pixels labeled 'skin'). It must be noticed that M1 must necessarily be used to exclude clusters that compose the concentric halo near the border between the bright skin and the black frame.

Examples of the global color cluster consisting by all the pixels labeled 'lesion' are in Figs. 10 and 11 respectively for the original image and for the one with hairs removed (for the future comparisons the pixels are in red).

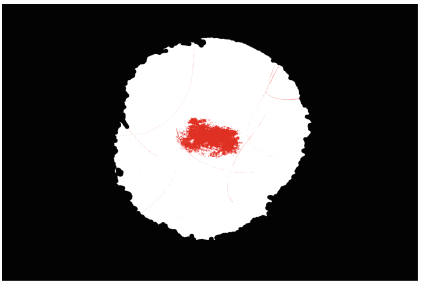

Fig. 10. Final global color cluster labeled 'lesion', no hairs removed (Color figure online)

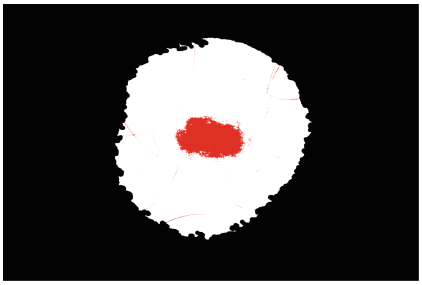

Fig. 11. Final global color cluster labeled 'lesion', hairs removed (Color figure online)

4 Experimental Setup and Results

To test which of the proposed detection pipelines reaches the best performance we will take from the dataset a small group of *hard images* where skin lesions differ from each other for size, colors, patterns, type (pigmented or not) and moreover for the presence of thick hairs at various sizes (Fig. 12).

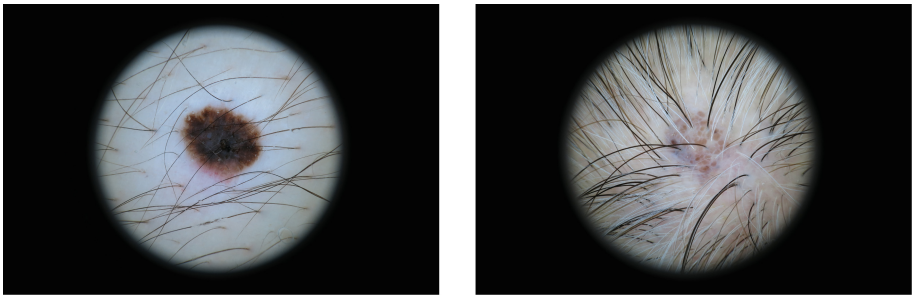

Fig. 12. Examples of hard dermoscopic images (Color figure online)

Each of the 17 chosen images are differentiated in 'original' version (Dataset 1) and 'hair removed' version (Dataset 2) for a total of 34 images. The experiment has a within-subjects design with two treatments and its structure is:

1. Dataset 1 - Otsu pipeline
2. Dataset 2 - Otsu pipeline
3. Dataset 1 - Color clustering pipeline
4. Dataset 2 - Color clustering pipeline

From a manual annotation we extract the green labeled area whose pixels must be considered as the *lesion ground truth* (Fig. 13). It is possible to consider, for each image in the datasets, the resulting labeled pixels (previously seen in red color) coming from the two detection pipelines and intersect them with the ground truth area to calculate measures like *Precision, Recall, f1-score* and *Accuracy*. To compute the four measures we consider a sort of *global goodness* for a detection pipeline, coming from the sum of the pixel of each classification, for each image of the dataset.

An example of the comparison between Figs. 13 and 6 for the Otsu pipeline is depicted in Fig. 14:

- true positives (TP): the pixel intersection (blue pixels)
- false positives (FP): red pixel that does not intersect the green ones
- false negatives (FN): green pixels that does not intersect the red ones
- true negatives (TN): pixels that are neither red nor green.

Fig. 13. Area derived from the manual annotation: ground truth (Color figure online)

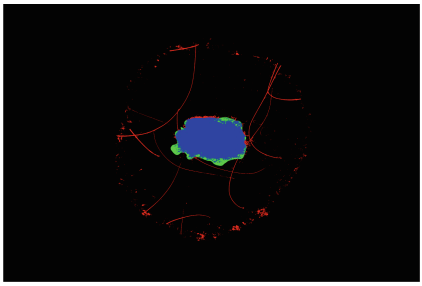

Fig. 14. Lesion detection comparison (blue pixels are the true positives) (Color figure online)

Experimental results are in Tables 1 and 2 for the Otsu pipeline while Tables 3 and 4 shows results for the Color clustering pipeline; it must be noticed that due to the peculiar image template and its size (pixel number) the *Accuracy* metric does not results significant, in fact the True Negatives (TN) represents most the large areas (as the black frame) that were considered and yet naturally excluded for pixel comparison during the detection pipelines.

Table 1. Results of Otsu pipeline on the original dataset

Otsu pipeline - Dataset 1				
Image	TP	FP	FN	TN
9001_2	217579	68714	33936	10335771
9012_2	70030	96724	2253	10486993
9013_3	364014	71399	14687	10205900
9017_2	87322	120510	5082	10443086
9030_2	296949	165099	3822	10190130
9051_2	7570	130537	219452	10298441
9067_2	106601	138285	10404	10400710
9084_2	52677	84071	3436	11859816
9108_2	214603	208246	1024	11576127
9110_2	176420	124390	12708	11686482
9112*a*_2	59915	47201	9617	11883267
9121_2	77964	41491	6187	11874358
9131_2	43994	109632	11788	11834586
9143_2	97002	113956	1150	11787892
9149_2	109552	61622	80655	11748171
9178*a*_2	645856	69300	71720	11213124
9190_2	818771	21707	82220	11077302
TOT.	3446819	1672884	570141	188902156
	Precision	*Recall*	*f1*	*Accuracy*
	0.67	0.86	0.75	0.99

Table 2. Results of Otsu pipeline on the hair removed dataset

Otsu pipeline - Dataset 2				
Image	TP	FP	FN	TN
9001_2	220376	43680	31139	10360805
9012_2	69964	80068	2319	10503649
9013_3	364010	66414	14691	10210885
9017_2	88283	106019	4121	10457577
9030_2	297509	73515	3262	10281714
9051_2	54816	129842	172206	10299136
9067_2	107926	116745	9079	10422250
9084_2	52832	57024	3281	11886863
9108_2	214525	158533	1102	11625840
9110_2	175242	92636	13886	11718236
9112*a*_2	59604	31391	9928	11899077
9121_2	77764	42813	6387	11873036
9131_2	44603	96025	11179	11848193
9143_2	96999	110616	1153	11791232
9149_2	114513	47094	75694	11762699
9178*a*_2	648944	36962	68632	11245462
9190_2	816557	8327	84434	11090682
TOT.	2476187	162813	1349042	190603958
	Precision	*Recall*	*f1*	*Accuracy*
	0.73	0.87	0.79	0.99

When comparing results between the two pipelines on each dataset we notice the detection improvement coming from the Color clustering especially in terms of the *Precision* metric, in fact it increases from 67% to 94% for the dataset 1 and from 73% to 94% for the second one. Also examining results between the two dataset for each pipeline we notice that all detection metrics always improve; only for the second dataset in Color Clustering pipeline, the *Precision* remains unchanged but *Recall* increases significantly): this demonstrates the need of hair removing treatment before each detection pipeline.

The *f1-score* metric provides the best interpretation of experimental results, in fact it increases in the Otsu pipeline (from 75% for dataset 1 to 79% for dataset 2) and in the Color clustering pipeline (from 79% for dataset 1 to 83% for dataset 2) and, as it is evident, it is always better for the second pipeline at the dataset equality. Finally, to explain these results it must be considered that, as previously explained, we considered images having specific characteristics and that, moreover, they represents only a part of the entire dermoscopic dataset.

Table 3. Results of Color clustering pipeline on the original dataset

Color clustering pipeline - Dataset 1				
Image	TP	FP	FN	TN
9001_2	126582	3625	111980	10413813
9012_2	38998	3192	25896	10587914
9013_3	299399	14928	64596	10277077
9017_2	0	0	84975	10571025
9030_2	116308	15577	170264	10353851
9051_2	0	0	215597	10440403
9067_2	88896	19563	19617	10527924
9084_2	0	0	50833	11949167
9108_2	190647	38220	10254	11760879
9110_2	133860	14066	44867	11807207
9112*a*_2	20617	1455	41831	11936097
9121_2	56131	739	20949	11922181
9131_2	0	0	49121	11950879
9143_2	86348	20269	3693	11889690
9149_2	82926	6518	97343	11813213
9178*a*_2	499044	17268	195287	11288401
9190_2	736431	7393	141939	11114237
TOT.	2476187	162813	1349042	190603958
	Precision	*Recall*	*f1-score*	*Accuracy*
	0.94	0.65	0.77	0.99

Table 4. Results of Color clustering pipeline on the hair removed dataset

Color clustering pipeline - Dataset 2				
Image	TP	FP	FN	TN
9001_2	183634	6196	54928	10411242
9012_2	57091	1600	7803	10589506
9013_3	299163	10266	64832	10281739
9017_2	59410	1423	25565	10569602
9030_2	278039	24196	8533	10345232
9051_2	0	0	215597	10440403
9067_2	82996	17204	25517	10530283
9084_2	30803	1022	20030	11948145
9108_2	198896	94849	2005	11704250
9110_2	156684	30981	22043	11790292
9112*a*_2	28484	120	33964	11937432
9121_2	56452	491	20628	11922429
9131_2	23825	3110	25296	11947769
9143_2	87097	21348	2944	11888611
9149_2	80332	268	99937	11819463
9178*a*_2	520811	3489	173520	11302180
9190_2	736402	1201	141968	11120429
TOT.	2880119	217764	945110	190549007
	Precision	*Recall*	*f1-score*	*Accuracy*
	0.93	0.75	0.83	0.99

5 Conclusions and Future Work

This work represents the second step (after the development and testing of the annotation tool) towards the development of a complex medical data management system used as an agent to support dermatologists in their decisions by exploiting all of the architecture modules while raw and structured data. The capacities must go from the images gathering and their annotation to the feature extraction and the analysis and visualization of their (meta)data, also exploiting Artificial Intelligence methods and CNN for advanced predictive performances.

Further tests and evaluations are needed, also considering the variety of lesion patterns and their stratification but the experiments presented shows encouraging results also with complex dermoscopic images, moreover the evaluation metrics proposed results adequate to verify and measure the best results: the color clustering technique featuring the pixel mask, the hair removing and the toleration areas reaches very positive performances.

References

1. DullRazor: a software approach to hair removal from images. Comput. Biolo. Med. **27**(6), 533–543 (1997)
2. Abbes, W., Sellami, D.: High-level features for automatic skin lesions neural network based classification. In: 2016 International Image Processing, Applications and Systems (IPAS), pp. 1–7, November 2016
3. Ahlberg, C., Williamson, C., Shneiderman, B.: Dynamic queries for information exploration: an implementation and evaluation. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 1992, pp. 619–626. ACM, New York, (1992). http://doi.acm.org/10.1145/142750.143054
4. Bakheet, S.: An SVM framework for malignant melanoma detection based on optimized HOG features. Computation **5**(1), 4 (2017)
5. Celebi, M.E., Wen, Q., Iyatomi, H., Shimizu, K., Zhou, H., Schaefer, G.: A state-of-the-art survey on lesion border detection in dermoscopy images. In: Dermoscopy Image Analysis, pp. 97–129. CRC Press (2015)
6. 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). doi:10.1007/978-3-319-24888-2_15
7. Codella, N.C.F., Nguyen, Q., Pankanti, S., Gutman, D., Helba, B., Halpern, A., Smith, J.R.: Deep learning ensembles for melanoma recognition in dermoscopy images. CoRR abs/1610.04662 (2016). http://arxiv.org/abs/1610.04662
8. Diepgen, T., Mahler, V.: The epidemiology of skin cancer. Br. J. Dermatol. **146**(s61), 1–6 (2002)
9. Emre Celebi, M., Wen, Q., Hwang, S., Iyatomi, H., Schaefer, G.: Lesion border detection in dermoscopy images using ensembles of thresholding methods. Skin Res. Technol. **19**(1), e252–e258 (2013)
10. Fan, H., Xie, F., Li, Y., Jiang, Z., Liu, J.: Automatic segmentation of dermoscopy images using saliency combined with Otsu threshold. Comput. Biol. Med. (2017). http://www.sciencedirect.com/science/article/pii/S001048251730080X
11. 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)
12. Jaleel, J.A., Salim, S., Aswin, R.: Artificial neural network based detection of skin cancer. IJAREEIE **1**, 200–205 (2012)
13. Liu, Z., Sun, J., Smith, M., Smith, L., Warr, R.: Unsupervised sub-segmentation for pigmented skin lesions. Skin Res. Technol. **18**(1), 77–87 (2012)
14. Maglogiannis, I., Doukas, C.N.: Overview of advanced computer vision systems for skin lesions characterization. IEEE Trans. Inf Technol. Biomed. **13**(5), 721–733 (2009)
15. Majtner, T., Yildirim-Yayilgan, S., Hardeberg, J.Y.: Combining deep learning and hand-crafted features for skin lesion classification. In: 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA), pp. 1–6, December 2016
16. Marín, C., Alférez, G.H., Córdova, J., González, V.: Detection of melanoma through image recognition and artificial neural networks. In: Jaffray, D.A. (ed.) World Congress on Medical Physics and Biomedical Engineering, Toronto, Canada, 7–12 June 2015. IP, vol. 51, pp. 832–835. Springer, Cham (2015). doi:10.1007/978-3-319-19387-8_204

17. Peruch, F., Bogo, F., Bonazza, M., Cappelleri, V.M., Peserico, E.: Simpler, faster, more accurate melanocytic lesion segmentation through MEDS. IEEE Trans. Biomed. Eng. **61**(2), 557–565 (2014)
18. Schaefer, G., Krawczyk, B., Celebi, M.E., Iyatomi, H.: An ensemble classification approach for melanoma diagnosis. Memet. Comput. **6**(4), 233–240 (2014)
19. Seidenari, S., Pellacani, G., Grana, C.: Early detection of melanoma by image analysis. In: Bioengineering of the Skin: Skin Imaging & Analysis, pp. 305–312. CRC Press (2006)
20. Singh, S., Stevenson, J., McGurty, D.: An evaluation of polaroid photographic imaging for cutaneous-lesion referrals to an outpatient clinic: a pilot study. Br. J. Plast. Surg. **54**(2), 140–143 (2001)
21. Wighton, P., Lee, T.K., Lui, H., McLean, D.I., Atkins, M.S.: Generalizing common tasks in automated skin lesion diagnosis. IEEE Trans. Inf. Technol. Biomed. **15**(4), 622–629 (2011)
22. Yuan, Y., Chao, M., Lo, Y.C.: Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. IEEE Trans. Med. Imaging **PP**(99), 1 (2017)
23. Zagrouba, E., Barhoumi, W.: A prelimary approach for the automated recognition of malignant melanoma. Image Anal. Stereol. **23**(2), 121–135 (2011)