PH^2 - A dermoscopic image database for research and benchmarking*

Teresa Mendonça[1]   Pedro M. Ferreira[1]   Jorge S. Marques[2]   André R. S. Marcal[1]   Jorge Rozeira[3]

Abstract— **The increasing incidence of melanoma has recently promoted the development of computer-aided diagnosis systems for the classification of dermoscopic images. Unfortunately, the performance of such systems cannot be compared since they are evaluated in different sets of images by their authors and there are no public databases available to perform a fair evaluation of multiple systems. In this paper, a dermoscopic image database, called PH^2, is presented. The PH^2 database includes the manual segmentation, the clinical diagnosis, and the identification of several dermoscopic structures, performed by expert dermatologists, in a set of 200 dermoscopic images. The PH^2 database will be made freely available for research and benchmarking purposes.**

I. INTRODUCTION

Skin cancer represents a serious public health problem because of its increasing incidence and subsequent mortality. Among skin cancers, malignant melanoma is by far the most deadly form. As the early detection of melanoma significantly increases the survival rate of the patient, several non-invasive imaging techniques, such as dermoscopy, have been developed to aid the screening process [1]. Dermoscopy involves the use of an optical instrument paired with a powerful lighting system, allowing the examination of skin lesions in a higher magnification. Therefore, dermoscopic images provide a more detailed view of the morphological structures and patterns as compared to the normally magnified images of the skin lesions [1], [2]. However, the visual interpretation and examination of dermoscopic images can be a time consuming task and, as shown by Kittler *et al.* [3], the diagnosis accuracy of dermoscopy significantly depends on the experience of the dermatologists. Several medical diagnosis procedures have been introduced in order to guide dermatologists, such as the pattern analysis, the ABCD rule, the 7-point checklist, and the Menzies method. Overall, in these methods several dermoscopic criteria (i.e. asymmetry, border, colors, differential structures) have to be assessed to produce the final clinical diagnosis. However, the diagnosis of skin lesions is still a challenging task even using these medical procedures mainly attending to the subjectivity of clinical interpretation and lacking of reproducibility [1], [2].

Several computer-aided diagnosis (CAD) systems of digital dermoscopic images have been proposed to assist the clinical evaluation of dermoscopic lesions. A CAD system has usually three stages, namely: image segmentation, feature extraction/selection, and lesion classification. The evaluation and validation of these automatic dermoscopic image analysis systems requires a reliable ground truth image database. Therefore, the availability and generation of a ground truth image database is of crucial importance, specially in the dermoscopy field, due to the non-existence of a public ground truth database of dermoscopic images. This task must be performed by expert dermatologists who have to manually segment and annotate each dermoscopic image.

Herein, a dermoscopic image database (PH^2), to be used as ground truth in the evaluation and validation of both segmentation and classification algorithms, is presented. The PH^2 database contains a total number of 200 melanocytic lesions, including 80 common nevi, 80 atypical nevi, and 40 melanomas. The rather small number of melanomas, compared with the other two types of melanocytic lesions, can be explained by two main reasons. First of all, the number of real cases of melanomas is actually much smaller than the other ones. In addition, as melanomas are usually not completely inserted in the image frame and present many image artifacts, they are not always suitable to be used as ground truth in the evaluation of CAD systems.

For each image in the database, the manual segmentation and the clinical diagnosis of the skin lesion as well as the identification of other important dermoscopic criteria are available. These dermoscopic criteria include the assessment of the lesion asymmetry, and also the identification of colors and several differential structures, such as pigment network, dots, globules, streaks, regression areas and blue-whitish veil. The PH^2 database will be made freely available (online) simultaneously with the EMBC 2013 conference. Meanwhile, a sample of the PH^2 database is already available for review in [4].

The size of the PH^2 database (200 images) might seems small, particularly when compared with a traditional machine learning ground truth database, which may have hundreds or thousands of annotated images. However, it is important to highlight that the annotation of dermoscopic images is not just a binary issue (benign or malign). The annotation of each image requires a large amount of time and effort, since several dermoscopic features have to be assessed to perform the lesion diagnosis. Moreover, the skin lesion and the color classes present in each image have to be manually segmented

[1]T. Mendonça, P. Ferreira and A. Marcal are with Faculdade de Ciências, Universidade do Porto, `tmendo at fc.up.pt`, `ferreira.pmmf at fc.up.pt` and `andre.marcal at fc.up.pt`

[2]J. Marques is with Instituto Superior Técnico and Institute for Systems and Robotics, Lisbon, `jsm at isr.ist.utl.pt`

[3]J. Rozeira is with Hospital Pedro Hispano, Matosinhos, Portugal, `jrozeira at me.com`

by expert clinicians.

Besides benchmarking of computer vision/machine learning algorithms, a database like PH^2 can be also used for medical training. For instance, dermatologist trainees can test their skills by comparing their own diagnosis and evaluation with the ground truth available in the PH^2 database.

The paper is organized in four sections including the Introduction (Section I). Section II gives an overview of the state of the art in CAD systems for melanoma diagnosis. In section III the PH^2 database is presented and described. Finally, conclusions and topics for future work are presented in Section IV.

II. AUTOMATIC MELANOMA DIAGNOSIS

Computer-aided diagnosis systems of dermoscopic images have been developed since 1987 [5], in order to support the clinical decision of dermatologists and easily trigger the detection of the highly suspicious cases. They can also be used as an additional tool by non-experienced clinicians to perform a preliminary evaluation and to improve the patient follow-up procedure [6]. These systems can be roughly divided in two main groups concerning the extraction of meaningful features from the dermoscopic images. One group tries to mimic medical diagnosis procedures and to automatically extract the same medical features (e.g., number of colors, symmetry, atypical differential structures). The other group is based on statistical pattern recognition and machine learning applied to standard image features (e.g., color and texture features).

An example of a diagnosis system inspired on clinical features can be found in [7]. As this automated diagnosis system is based on the 7-point checklist, the authors describe methods to detect atypical network, blue-whitish veil, atypical vascular pattern, irregular streaks, regression areas, irregular pigmentation and irregular dots/globules. The system as a whole achieved a sensibility (SE) of 83% and a specificity (SP) of 76% in a database of 287 images. Gola *et al*. [8] propose an automated system for melanoma diagnosis, mainly based on the pattern analysis, in which both globular and reticular patterns are detected. In [9], another system for automatic detection of melanoma is proposed. This system is based on the dermoscopic criteria evaluated in the ABCD rule of dermoscopy, including the asymmetry, border irregularity, color and diameter.

The computer-aided diagnosis based on the classic machine learning methodology has been explored to a much larger extent [10], [11], [12], [13]. Ganster *et al*. [10] suggest a system in which a large number of color, texture and border features are extracted and then selected as the input to a k-nearest neighbors (kNN) classifier. Their system achieved a SE of 87% with a SP of 92%. In [11], border and color-based features are extracted to be used in a classification module based on artificial neural networks (ANN). The system proposed in [12] uses a set of shape, color, and texture related features and a support vector machine (SVM) to classify the skin lesion. They obtained a SE of 93% and a SP of 92% in a database of 564 images. Iyatomia *et al*.

[13] propose a melanoma diagnosis system that is already available in the internet for public use. This system uses a set of features to characterize the color, symmetry, border, and texture of a skin lesion, and a ANN classifier to provide the diagnosis.

Although the automatic diagnosis approaches demonstrate promising results, the evaluation of their performance is still a problem since each research group uses its own image database. Therefore, at this point, the creation of a dermoscopic image database that could be used as ground truth by the research community is a fundamental issue. The database presented in this paper attempts to address this problem.

III. PH^2 DATABASE

The PH^2 database was built up through a joint research collaboration between the Universidade do Porto, Técnico Lisboa, and the Dermatology service of Hospital Pedro Hispano in Matosinhos, Portugal. The dermoscopic images were obtained under the same conditions through Tuebinger Mole Analyzer system using a magnification of $20\times$. They are 8-bit RGB color images with a resolution of 768×560 pixels.

This image database contains a total of 200 dermoscopic images, containing 80 common nevi, 80 atypical nevi, and 40 melanomas. All dermoscopic images are either from the skin type II or III, according to the Fitzpatrick skin type classification scale [14]. Therefore, the skin colours represented in the PH^2 database may vary from white to cream white. As illustrated in Fig.1, the images of the database were carefully selected taking into account their quality, resolution and dermoscopic features. Every image was evaluated by an expert dermatologist with regard to the following parameters:

- Manual segmentation of the skin lesion;
- Clinical and histological (when available) diagnosis;
- Dermoscopic criteria (Asymmetry; Colors; Pigment network; Dots/Globules; Streaks; Regression areas; Blue-whitish veil).
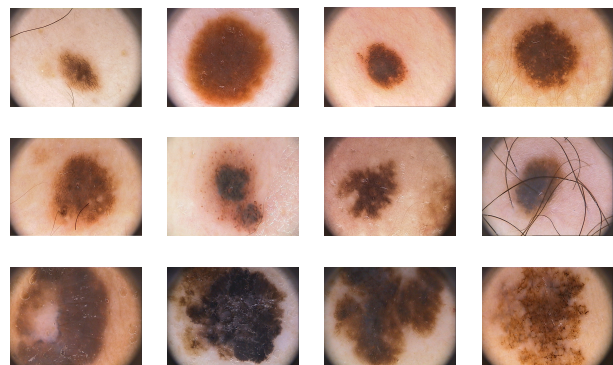

Fig. 1: An illustrative collection of images from PH^2 database, including common nevi (1st row), atypical nevi (2nd row) and melanomas (3rd row).
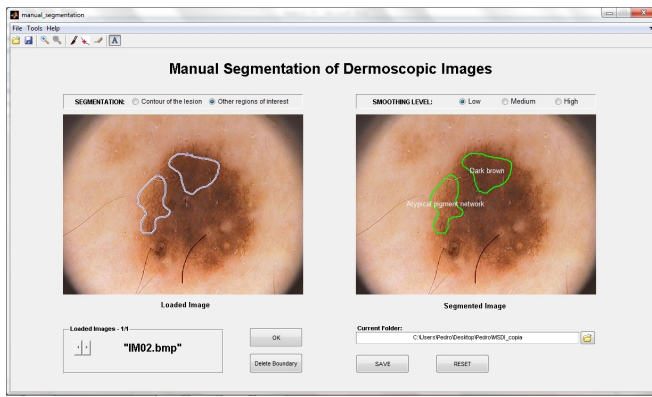
Fig. 2: DerMAT interface for the segmentation and labeling of multiple regions of interest.

Fig. 3: Manual segmentation of three melanocytic lesions: common nevus (left), atypical nevus (middle) and melanoma (right).

Dermatologists performed the manual segmentation and annotation of the images using a customized annotation tool for dermoscopic images, called DerMAT [15]. As an example, Fig.2 shows the manual segmentation and annotation of two regions of interest using the DerMAT software.

A. Manual segmentation of the skin lesion

The manual segmentation of the skin lesion, performed by expert dermatologists, is an essential information for the evaluation of the segmentation step of a CAD system.

In this database, the manual segmentation of each image is available as a binary mask, in which pixels with intensity value of 1 correspond to the segmented lesion, while pixels with value 0 correspond to the background. This binary mask has the same size of the original image and, hence, it can be easily used to extract the boundary coordinates of the lesion. Figure 3 presents examples of three dermoscopic images and the corresponding ground truth (manual) segmentations.

B. Clinical diagnosis

The melanocytic lesions can be divided in two main groups concerning their nature: benign lesions (which include common and atypical nevus) and malignant lesions (or melanomas). Therefore, each image of the database is classified into common nevus, atypical nevus, or melanoma (Fig.3). The histological diagnosis is only available for some of the images, since the histological test is just performed for those lesions considered highly suspicious by dermatologists.

C. Dermoscopic criteria

The set of dermoscopic features that is available in the PH^2 database corresponds to those features that the dematologists of Hospital Pedro Hispano consider more relevant to perform a clinical diagnosis. This set of features includes the majority of the dermoscopic features that have to be assessed in the most widely used medical diagnosis procedures, such as the ABCD rule, the 7-point checklist and the Menzies method. All of these features as well as their evaluation process are described below.
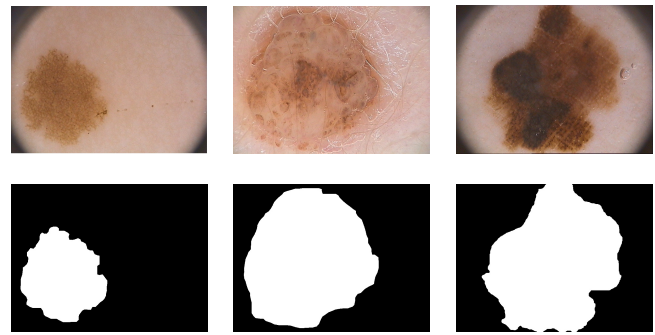
1) Asymmetry:

The asymmetry is one of the most important features for diagnosing a melanocytic lesion. For instance, it is the largest weight factor in the ABCD rule of dermoscopy.

In this database, the lesion asymmetry was evaluated by the clinician according to the ABCD rule. Therefore, the asymmetry of a lesion is assessed regarding its contour, colors, and structures distribution simultaneously. Moreover, there are three possible labels for this parameter: 0 for fully symmetric lesions; 1 for asymmetric lesions with respect to one axis; and 2 for asymmetric lesions with respect to two axes.

2) Colors:

Overall, six different colors are taken into account during the diagnosis of a melanocytic lesion. The set of color classes comprises the white, red, light-brown, dark-brown, blue-gray, and black [1].

Each image of the database was evaluated by a dermatologist in order to identify the presence, as well as the location, of the six color classes. The location of each color in an image was recorded as a binary mask, manually segmented by the dermatologist. An example is presented in Fig.4, where two color classes were identified.

3) Pigment network:

The pigment network is a grid-like network consisting of pigmented lines (brown or black) and hypopigmented holes [1]. This structure has a crucial role in the distinction between melanocytic and non-melanocytic lesions as well as

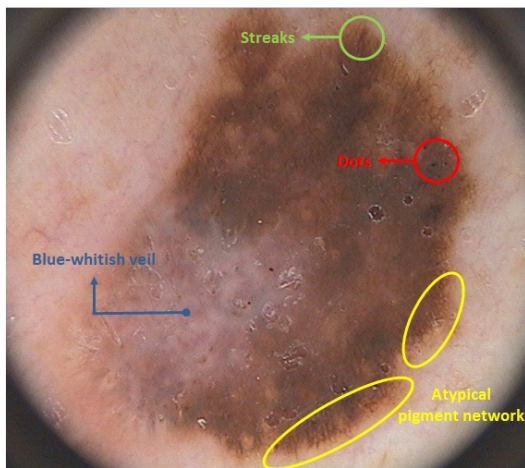Fig. 4: Original image (left), blue-gray (middle) and dark-brown region (right).

Fig. 5: Dermoscopic features identification.

in melanoma diagnosis. The pigment network structure was visually evaluated by the dermatologist, in each image of the database, and classified as typical or atypical. Samples of atypical pigmented network can be found in Fig.5.

4) Dots/Globules:

As illustrated in Fig.5, dots/globules are spherical or oval, variously sized, black, brown or gray structures (dots are usually smaller than globules). The presence of these dermoscopic structures is also particularly useful for the distinction between melanocytic and non-melanocytic lesions [1].

These structures were visually evaluated by dermatologists, and categorically classified as present or absent in each image of the PH^2 database. When dots/globules are present in a given lesion, these structures are further classified as regular or irregular concerning their distribution in the lesion.

5) Streaks:

Streaks are finger-like projections of the pigment network from the periphery of the lesion. Instead of both pigment network and dots/globules, the presence of streaks in a skin lesion is by itself a sign of malignancy [1]. Therefore, these structures are just classified as present or absent in each image of the database. Fig.5 illustrates the presence of a streak in a skin lesion (identified on the upper area).

6) Regression areas:

Regression areas are defined as white, scar-like depigmentation often combined with pepperlike regions (speckled blue-gray granules) [1]. In the PH^2 database, this parameter is classified in two main groups (present or absent) concerning its presence in the skin lesion.

7) Blue-whitish veil:

The blue-whitish veil can be defined as a confluent, opaque, irregular blue pigmentation with an overlying, white, ground-glass haze. Its presence is a strong malignancy indicator [1]. This dermoscopic structure is categorically labeled

as present or absent, in each image of the database. The presence of the blue-whitish veil in a melanocytic lesion is illustrated in Fig.5.

IV. CONCLUSIONS

This paper presents a database of dermoscopic images, PH^2, acquired at Pedro Hispano Hospital. This database includes medical annotation of all the images namely medical segmentation of the lesion, clinical diagnosis and dermoscopic criteria (asymmetry, colors and the presence of typical and atypical differential structures). The database will be freely available through internet to all researchers for benchmarking purposes after EMBC 2013 conference. This first version will be released with 200 annotated images, and it will be gradually incremented as new clinical cases are collected from the hospital.

REFERENCES

[1] "Dermoscopy tutorial," http://www.dermoscopy.org.

[2] G. C. do Carmo and M. R. e Silva, "Dermoscopy: basic concepts," *Int J Dermatol.*, vol. 47, no. 7, pp. 712–719, July 2008.

[3] H. Kittler, H. Pehamberger, K. Wolff, and M. Binder, "Diagnostic accuracy of dermoscopy," *The Lancet Oncology*, vol. 3, no. 3, pp. 159–65, March 2002.

[4] "Addi project," http://www2.fc.up.pt/addi.

[5] N. Cascinelli, M. Ferrario, T. Tonelli, and E. Leo, "A possible new tool for clinical diagnosis of melanoma: the computer," *Journal of the American Academy of Dermatology*, vol. 16, no. 2, pp. 361–367, 1987.

[6] K. Korotkov and R. Garcia, "Computerized analysis of pigmented skin lesions: A review," *Artificial Intelligence in Medicine*, 2012.

[7] G. D. Leo, A. Paolillo, P. Sommella, G. Fabbrocini, and O. Rescigno, "A software tool for the diagnosis of melanomas," in *IEEE Instrumentation and Measurement Technology Conference*, 2010, pp. 886–891.

[8] A. I. Gola, B. G. Zapirain, A. M. Zorrilla, and I. R. Oleagordia, "Automated diagnosis of melanomas based on globular and reticular pattern recognition algorithms for epiluminescence image," in *European Signal Processing Conference*, 2010.

[9] J. J.-Korjakowska, "Automatic detection of melanomas: An application based on the abcd criteria," in *Information Technologies in Biomedicine.* Springer Berlin Heidelberg, 2012, vol. 7339, pp. 67–76.

[10] H. Ganster, A. Pinz, R. Röhrer, and E. W. *et al.*, "Automated melanoma recognition," *Medical Imaging, IEEE Transactions on*, vol. 20, no. 3, pp. 233–239, 2001.

[11] I. Maglogiannis, S. Pavlopoulos, and D. Koutsouris, "An integrated computer supported acquisition, handling, and characterization system for pigmented skin lesions in dermatological images," *IEEE Transactions on Information Technology in Biomedicine*, pp. 86–98, 2005.

[12] M. E. Celebi, H. A. Kingravi, B. Uddin, and H. I. *et al.*, "A methodological approach to the classification of dermoscopy images," *Computerized Medical Imaging and Graphics*, pp. 362–373, 2007.

[13] H. Iyatomi, H. Oka, M. E. Celebi, and M. H. *et al.*, "An improved internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm," *Computerized Medical Imaging and Graphics*, vol. 32, no. 7, pp. 566–579, 2008.

[14] S. Sachdeva, "Fitzpatrick skin typing: Applications in dermatology," *Indian J Dermatol Venereol Leprol.*, vol. 75, no. 1, pp. 93–96, 2009.

[15] P. M. Ferreira, T. Mendonça, J. Rozeira, and P. Rocha, "An annotation tool for dermoscopic image segmentation," in *Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications.* ACM, 2012, pp. 5:1–5:6.