Pigmented skin lesions classification using dermatoscopic images

Germán Capdehourat[1], Andrés Corez[1], Anabella Bazzano[2], and Pablo Musé[1]

[1] Departamento de Procesamiento de Señales, Instituto de Ingeniería Eléctrica, Facultad de Ingeniería, Universidad de la República, Uruguay
[2] Unidad de Lesiones Pigmentadas, Cátedra de Dermatología, Hospital de Clínicas, Facultad de Medicina, Universidad de la República, Uruguay

Abstract. In this paper we propose a machine learning approach to classify melanocytic lesions in malignant and benign from dermatoscopic images. The image database is composed of 433 benign lesions and 80 malignant melanoma. After an image pre-processing stage that includes hair removal filtering, each image is automatically segmented using well known image segmentation algorithms. Then, each lesion is characterized by a feature vector that contains shape, color and texture information, as well as local and global parameters that try to reflect structures used in medical diagnosis. The learning and classification stage is performed using AdaBoost.M1 with C4.5 decision trees. For the automatically segmented database, classification delivered a false positive rate of 8.75% for a sensitivity of 95%. The same classification procedure applied to manually segmented images by an experienced dermatologist yielded a false positive rate of 4.62% for a sensitivity of 95%.

1 Introduction

The incidence of melanoma in the general population is increasing worldwide. It is estimated that by the end of this decade, four million new melanomas will be diagnosed in the world, causing the death of half million people. If early diagnosed and treated, the mean life expectancy of these individuals would have been enlarged by at least 25 years. Because advanced cutaneous melanoma is still incurable, early detection, by means of accurate screening, is an important step toward mortality reduction. Detection of thin malignant melanoma is the most effective way to avoid mortality related to this disease.

Dermoscopy is a noninvasive in vivo technique that assists the clinician in melanoma detection in its early stage. Images are acquired using epiluminescence light microscopy, that magnifies lesions and enables examination down to the dermo-epidermal junction. This permits to visualize new morphologic features and in most cases facilitates early diagnosis. However, evaluation of the many morphologic characteristics is often extremely complex and subjective [1].

Advances in objective dermatology diagnosis were obtained in 1994 with the introduction of the ABCD rule [2]. The ABCD rule specifies a list of visual features associated to malignant lesions (Asymmetry, Border irregularity,

Color irregularity and presence of Dermoscopic structures), from which a score is computed. This methodology provided clinicians with a useful quantitative criterion, but it did not prove efficient enough for clinically doubtful lesions (CDL). The main reason for this is the difficulty in visually characterizing the lesions' features. Setting an adequate decision threshold for the score is also a difficult problem; by now it has been fixed based in several years of clinical experience. Many authors claim that these thresholds may lead to high rates of false diagnoses [3]. An alternative algorithm for melanocytic lesion diagnosis is the 7-points checklist [4]. This algorithm consists of analyzing the presence of the seven most important color or geometric stuctures that characterize malignant melanoma (blue whitish veil, atypical pigment network, irregular streaks, etc.).

The computerized analysis of dermatoscopic images can be an extremely useful tool to measure and detect sets of features from which dermatologists make their diagnosis. It can also be helpful for primary screening campaigns, increasing the possibility of early diagnosis of melanoma. Currently there is no commercial software for massive use in clinical practice. Our ultimate goal is to develop software for the recognition of early-stage melanomas, based on images obtained by digital dermoscopy. This would enable unsupervised classification of melanocytic lesions, assigning a confidence index for each classification. The result of such classification procedure will separate the "screened" lesions in two groups. The first group corresponds to lesions that were classified with high enough confidence level, while the second one corresponds to those lesions for which the confidence level is low and consequently, requires subsequent inspection by an experienced dermatologist. In this sense, the classification technique is actually a semi-automated method.

The paper is organized as follows. In Section 2 we present a brief overview of previous related work. In Section 3 we describe the composition of our database of dermatoscopic images, and in Section 4 we present our approach to melanocytic lesions classification. Results and performance are presented and discussed in Section 5. We conclude in Section 6.

2 Computerized analysis of dermoscopic images: state of the art

Computer aided image analysis in skin lesion diagnosis is a relatively new research field. While the first related work in the medical literature seems to date back to 1987 [5], its contribution was limited since by that time computer vision and machine learning were both emerging fields (the first edge detectors where starting to appear). One of the first significant contributions from the image processing community was reported in [6]. In this work, the authors propose a classical machine learning approach for dermatoscopic image classification. The first stage is automatic color-based lesion segmentation.Then, over a hundred features are extracted from the image (shape and color, and gradient distribution in the neighborhood of the lesion boundary). Feature selection was obtained using sequential forward and sequential backward floating selection. Classifica-

tion experiments, performed with a 24-NN classifier, delivered a sensitivity of 77% with a specificity of 84%.

To our knowledge, up to now the best results in automated melanocytic lesion classification where obtained by Celebi *et al.* [7]. See this reference for a complete summary of the results obtained by key studies from 2001 onwards, along with their database sizes. As in [6], the proposed approach is a classic machine learning methodology. After an Otsu-based image segmentation, a set global features are computed (area, aspect ratio, asymmetry and compactness). Local color and texture features are computed after dividing the lesion in three regions: inner region, inner border (an inner band delimited by the lesion boundary) and outer border (an outer band delimited by the lesion boundary). Feature selection is performed using ReliefF [8] and CFS algorithms [9]. Finally, the feature vectors are classified into malignant and benign using SVM with model selection [10]. Performance evaluation gave a specificity of 92.34% and a sensitivity of 93.33%.

3 Database composition

Our database is composed of 513 images of melanocytic lesions: 433 benign lesions and 80 malignant melanoma. Among the set of benign lesions, over a hundred correspond to dysplastic melanocytic nevi. It is important to note that in general these kind of lesions are the benign lesions that are visually the most alike to malignant melanoma; many of them are clinically doubtful for experienced dermatologists.This composition was based on the existence of dermatoscopic and histopathologic studies, which were used as ground truth for the classification procedure. Actually, the original database was larger, but some images were discarded for the following reasons: the images do not capture the whole lesion, poor image quality or excessive presence of hair.

Every image in this database has been manually segmented by a dermatologist, who also provided dermatoscopic diagnosis based on the ABCD rule and the 7-points checklist. This enables performance evaluation for both segmentation and features' measurements.

4 Dermoscopic images classification: proposed approach

Our approach follows a typical machine learning methodology. In the first stage, we tackle image processing problems such as image filtering, restoration and automatic segmentation to isolate the lesion's area. The second stage consists of extracting features from the image for further lesion classification into malignant or benign. Features are inspired by the same elements that dermatologists use for lesion diagnosis. Once lesions' features have been extracted, labeled lesions are used to train a meta-classifier obtained using boosting based on decision trees. Classification errors and ROC curves are obtained by means of cross validation. In this section we give details of each of these stages.

4.1   Preprocessing and hair removal

Lesion segmentation in the presence of hair is usually doomed to failure. Thus, previous application of a hair removal filter is unavoidable. Automatic hair removal requires hair detection and image inpainting. We used Dullrazor [11], a well known algorithm for hair removal. This algorithm identifies the image segments that approximate the structure of the hair, and then the regions that contain these segments are interpolated using the information of the surrounding pixels. A typical result is shown in Figure 1(a)(b). For the inpainting part, more sophisticated techniques were also explored, with similar results.

4.2   Segmentation

Segmentation of melanocytic lesions can be an extremely hard problem. Besides the presence of hair, many lesions present diffuse borders, that can be difficult to determine even for dermatologists. Several methods of image segmentation were explored, based on edge detection and on region information. In general it is appropriate to combine different features (texture, edges, color) for better results. Methods combining these sources of information were also studied. Among the variational methods family, we considered Otsu using color norm instead of grey level [12], Mumford-Shah [13], Geodesic Active Contours and Geodesic Active Regions [14]. We explored also several methods based on the topographic map, using both boundary and color and texture region information [15, 16]. We are currently investigating spectral clustering – graph based approaches.

Overall, none of the methods outperformed the others. We decided to use the color-based Otsu method for it is simpler and significantly faster. Of course, there are pathological cases in which it fails, and sometimes one of the others provides satisfactory results. This suggests that a software for clinical use should propose the choice of a few candidate segmentations to the user in case they differ.

4.3   Feature extraction

A set of global measurements of shape (aspect ratio, symmetry, compacity, etc.) and border irregularity were computed from each lesion. More localized features of texture and color distribution were also extracted. Previous to their extraction, each lesion is decomposed into three sub-regions: the interior and the outer and inner border (Figure 1). For each of these regions, the color features consist of some statistics of its distribution (mean and variances per channel in RGB and HSV spaces), and the texture features based on Gabor filters capture information of local contrast, correlation, heterogeneity and energy. For each lesion, a total number of 57 features are extracted.

Note that information concerning the presence or absence of several geometric patterns that are relevant to the 7 points checklist is not included in the feature vectors. This requires the detection of these structures, which is not a trivial task, what explains why they are not included in any previous work, either. We are currently investigating these detection problems, for we are confident that the capability of detecting this structures will boost our method performance.
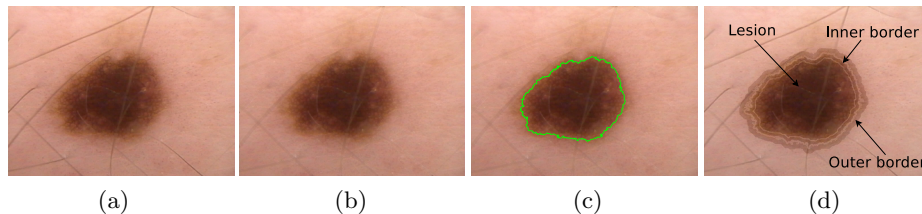
Fig. 1. (a) Original lesion. (b) Result of the hair removal filter. (c) Color-based Otsu segmentation. (d) Definition of the three regions used for feature extraction.

4.4 Classification

The goal of this stage is to classify the feature vectors in two classes: malignant and benign. A classification technique that prove very successful in our experiments consist of performing decision trees combination *via* adaptive boosting. Boosting exploits the inherent instability in learning algorithms by combining multiple models, in a way that models complement one another. This is achieved by assigning weights to the training data, and modifying them after each classifier by increasing the weight of misclassified samples, and decreasing these of correctly classified ones. Hence, after each iteration, a new classifier is forced to focus on classifying the hard samples correctly. The algorithm finishes after a user-defined number of T iterations, that generates a set of T classifiers. To each of them, a weight that increases with its performance is associated. Classification of new unlabeled data is performed by a weighted vote of the T classifiers.

The algorithms we considered for the classification framework are C4.5 decision trees [17], and AdaBoost.M1 [18], using Weka's implementations. In order to deal with class imbalance, we applied a widely used synthetic over-sampling technique (SMOTE [19]) to the minority class.

5 Results

Performance evaluation was conducted using 10 times - 10 fold cross-validation. To assess the impact of the learning and classification method, we compared our results with SVM with model selection (preceeded by ReliefF feature selection). As in [7], a RBF kernel was used, and optimal parameters (the weight that controls model complexity and the RBF parameter) were obtained by grid search optimization with 10 fold cross-validation. Classification performance was also estimated using 10 times - 10 fold cross-validation.

The same experiments were repeated, replacing automatic segmentation by manual segmentation by a dermatologist. This was carried on to assess the influence of automatic segmentation errors.

The left plot in Figure 2 shows the overall system performance using automatic segmentation, for both learning strategies. The right plot shows the results for the manually segmented images. In both cases, the AdaBoost/C4.5 method
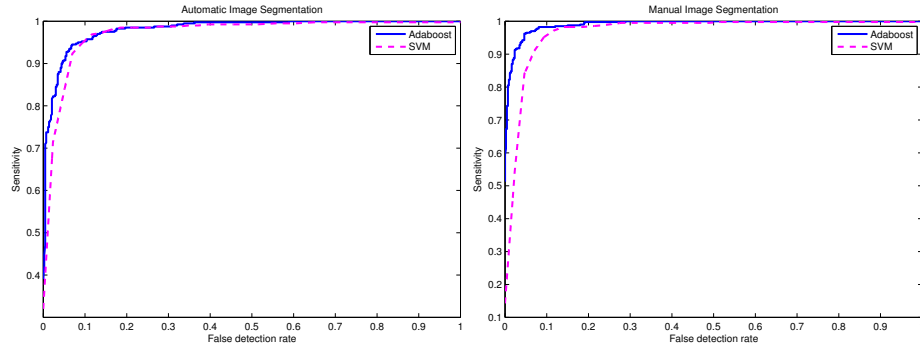
Fig. 2. Left: ROC curves for the AdaBoost/C4.5 and SVM approaches for automatically segmented (left) and manually segmented (right) images. See text for details.

outperformed the SVM-based approach. Table 1 shows performance indicators for the four experiments.

While the SVM approach using manually or automatically segmented images yielded essentially the same performance, the performance of Adaboost/C4.5 classification of manually segmented images was significantly higher than for the automatically segmented ones. Note that the results we obtained with SVM are slightly better than those reported by Celebi *et al.* [7] (false positive rate of 14% for 95% sensitivity and AUC of 0.966). Our AdaBoost/C4.5 approach shows even higher performance. Note that since the database used by Celebi *et al.* is very similar to ours in size and composition (476 benign lesions and 88 malignant melanoma), this performance comparison makes sense, but only up to a certain point.

Method	FPR for 95% sensitivity	Area under ROC
Automatic segmentation, AdaBoost - C4.5	8.75 %	0.981
Automatic segmentation, SVM	9.52 %	0.963
Manual segmentation, AdaBoost - C4.5	4.62%	0.991
Manual segmentation, SVM	9.23 %	0.966

Table 1. Performance indicators for the ROC curves in Figure 2.

Figure 3 shows the five misclassified patterns that correspond to lesion images in the database, for the AdaBoost/C4.5 classification of automatically segmented lesions. Among these lesions, all false positives were dysplastic melanocytic nevi, actually suspicious lesions according to the ABCD rule (CDL scores range from 4.75 to 5.45). Moreover, note that the rightmost one qualifies as melanoma according to the 7-points checklist algorithm (larger or equal than 3 corresponds to malignant melanoma). Concerning the false negatives, posterior inspection by an expert dermatologist revealed subjective overestimation of their scores, since the lesions corresponded to a patient with clinical history of melanoma.
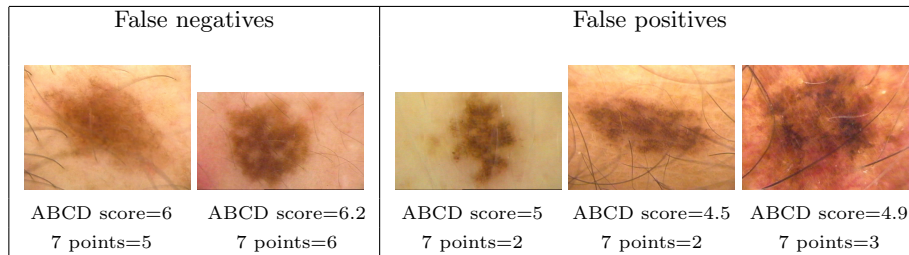
False negatives		False positives		
ABCD score=6	ABCD score=6.2	ABCD score=5	ABCD score=4.5	ABCD score=4.9
7 points=5	7 points=6	7 points=2	7 points=2	7 points=3

Fig. 3. All misclassified patterns corresponding to lesion images. Color-based Otsu segmentation was used. See text for details.

6 Conclusions and future work

In this work we presented a machine learning approach to classify melanocytic lesions from dermatoscopic images. The learning and classification stage is performed using AdaBoost.M1 with C4.5 decision trees. Using automatically segmented images, we obtained a false positive rate of 8.75% for a sensitivity of 95%, and an AUC of 0.981. These results are promising and seem to be superior than those reported in the literature. However, performance evaluation is delicate because all reported results were obtained using different databases. At this point, construction of a large database of dermatoscopic images that could be used as reference testbed appears to be a fundamental issue.

Concerning our algorithm, to further improve its performance, methods to detect a larger number of geometry or texture based structures, similar to those used in the 7 points checklist, should be developed. Because of their strong discriminative power, we are confident that the inclusion of these patterns' information in the features vectors will boost the classification results. This is ongoing research and hopefully will be implemented in future versions. It seems also, from the comparison of the results obtained from manually segmented lesions (FPR of 4.62% for a sensitivity of 95%), that errors in automatic segmentation have an important impact and should be reduced. As we pointed out, this is a hard problem since many melanocytic lesions show highly diffuse contours. Note, however, that nothing prevents us to manually segment the training database, and to propose to the user, for each new lesion, the choice of candidate segmentations.

Another interesting related line of research is the characterization of the discriminative power of the considered features. This can be obtained by means of automatic feature selection strategies like the ones that were mentioned here. A rigorous study of this topic, complemented with the comparison of the weights assigned to visual features in the ABCD and other clinical diagnosis rules, may yield useful recommendations to dermatologist for their medical practice.

References

1. Rubegni, P., Burroni, M., Dell'eva, G., Andreassi, L.: Digital dermoscopy analysis for automated diagnosis of pigmented skin lesion. Clinics in Dermatology **20**(3) (2002) 309–312
2. Nachbar, F., Stolz, W., Merkle, T., Cognetta, A., Vogt, T., Landthaler, M., Bilek, P., Braun-Falco, O., Plewig, G.: The ABCD rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions. Journal of the American Academy of Dermatology **30**(4) (1994) 551–559
3. Lorentzen, H., Weismann, K., Kenet, R., Secher, L., Larsen, F.: Comparison of dermatoscopic abcd rule and risk stratification in the diagnosis of malignant melanoma. Acta Derm Venereol. **80**(2) (2000) 122–126
4. Johr, R.H.: Dermoscopy: alternative melanocytic algorithms?the abcd rule of dermatoscopy, menzies scoring method, and 7-point checklist. Clinics in Dermatology **20**(3) (2002) 240–247
5. Cascinelli, N., Ferrario, M., Tonelli, T., Leo, E.: A possible new tool for clinical diagnosis of melanoma: The computer. Journal of the American Academy of Dermatology **16**(2) (1987) 361–367
6. Ganster, H., Pinz, A., Rhrer, R., Wildling, E., Binder, M., Kittler, H.: Automated melanoma recognition. IEEE Transactions on Medical Imaging **20** (2001) 233–239
7. Celebi, M.E., Kingravi, H.A., Uddin, B., Iyatomi, H., Aslandogan, Y.A., Stoecker, W.V., Moss, R.H.: A methodological approach to the classification of dermoscopy images. Comput Med Imaging Graph **31**(6) (2007) 362–373
8. Robnik-Šikonja, M., Kononenko, I.: Theoretical and empirical analysis of relieff and rrelieff. Mach. Learn. **53**(1-2) (2003) 23–69
9. Hall, M.A.: Correlation-based feature selection for discrete and numeric class machine learning. In: ICML '00: Proceedings of the 7th International Conference on Machine Learning, San Francisco, CA, USA (2000) 359–366
10. Schlkopf, B., Smola, A.J.: Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. The MIT Press (2001)
11. Lee, T., Ng, V., Gallagher, R., Coldman, A.: Dullrazor: A software approach to hair removal from images. Computers in Biology and Medicine **27**(11) (1997) 533–543
12. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man and Cybernetics **9**(1) (1979) 62–66
13. Koepfler, G., Lopez, C., Morel, J.M.: A multiscale algorithm for image segmentation by variational method. SIAM J. Numer. Anal. **31**(1) (1994) 282–299
14. Paragios, N., Deriche, R.: Geodesic active regions: A new framework to deal with frame partition problems in computer vision. Journal of Visual Communication and Image Representation **13** (2002) 249–268
15. Cao, F., Musé, P., Sur, F.: Extracting meaningful curves from images. Journal of Mathematical Imaging and Vision **22**(2-3) (2005) 159–181
16. Cardelino, J., Randall, G., Bertalmio, M., Caselles, V.: Region based segmentation using the tree of shapes. In: IEEE International Conference on Image Processing, Proceedings. (2006)
17. Quinlan, R.J.: C4.5: Programs for Machine Learning (Morgan Kaufmann Series in Machine Learning). Morgan Kaufmann (1993)
18. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. **55**(1) (1997) 119–139
19. Nitesh V. Chawla, N., Bowyer, K., Hall, L., Kegelmeyer, W.: Smote: Synthetic minority over-sampling technique. J. Artif. Intell. Res. (JAIR) **16** (2002) 321–357