The Impact of the Observation of Predictive Features on the Diagnosis of Pigmented Skin Lesions and the therapeutic decision

Y. Wazaefi, Arthur Tenenhaus, Alex Nkengne, Jean-Francois Horn, Alain Giron, Sébastien Paris, Bernard Fertil

HAL Id: hal-00756094

https://centralesupelec.hal.science/hal-00756094

Submitted on 22 Nov 2012

The impact of the observation of predictive features on the diagnosis of pigmented skin lesions and the therapeutic decision

Y. Wazaefi[1], A. Tenenhaus[2], A. Nkengne[3], J.F. Horn[4], A. Giron[4], S. Paris[1] and B. Fertil[1]

[1]LSIS, Aix-Marseille University, Marseille, France
[2] Department of Signal Processing and Electronic Systems, ESE Plateau de Moulon, Gif-sur-Yvette, France
[3]Johnson and Johnson Consumer, Skin Care Research Institute, Issy-les-moulineaux, France
[4]INSERM, Pierre et Marie Curie-Paris University, Paris, France

Abstract - *In this paper, we study the relationship between diagnosis and therapeutic decision on the one hand and the observations of the presence of ABCD features and some additional dermoscopic features of pigmented skin tumors on the other hand. The image database was composed of 227 images of pigmented skin lesions. Five senior dermatologists were asked for their expertise about these images. They gave their opinion about the presence of ABCD and dermoscopic features, their diagnosis and their therapeutic decision. The performances of dermatologists were evaluated in terms of their ability to diagnose melanoma by building statistical decision models from their observations of predictive features. Models allowed observing to what extent dermatologists ground their diagnosis on the malignancy features they detected. It appeared that a high variability of behavior among dermatologists is observed, concerning both the detection of features and the role of features for the elaboration of diagnosis.*

Keywords: ABCD features, melanoma diagnosis, decision model, Roc curve, Medicine Data Mining.

1   Introduction

As the survival rate of malignant melanoma depends on its thickness, diagnosis of malignant melanoma at an early stage could reduce the risk of mortality and increase the chance of prognosis considerably. The accuracy of the clinical diagnosis of melanoma with the unaided eye is only about 60%. Dermoscopy is a non-invasive in vivo technique for the microscopic examination of pigmented skin lesions, has the potential to improve the diagnostic accuracy [1]. Advances in objective dermatology diagnosis were obtained in 1994 with the introduction of the ABCD rule [2-3]. The ABCD rule specifies a list of visual features associated to malignant lesions (Asymmetry, Border irregularity, Color irregularity and Differential structure, i.e. size and number of structural features), from which a score is computed [4]. This methodology provided clinicians with a useful quantitative criterion, but it did not prove efficient enough

for clinically doubtful lesions (CDL) essentially because ABCD features are difficult to characterize in those situations [5].

According to dermatologists' 'rules of good clinical practice', the diagnosis and associated therapeutic decision for black skin tumors is a multi-step procedure. The first step consists in detecting malignancy features (ABCD rule, 7-points checklist [6], etc.). In the second step, dermatologists combine these features according to their capacity in predicting malignancy. Stolz et al. has formulated a mathematical implementation of the ABCD rule [4]. Given that feature A may get a score varying from 0 to 2, feature B a score varying from 0 to 8, feature C a score varying from 1 to 6 and feature D a score varying from 1 to 5, a decision score (TDS) may be obtained by a linear combination of the features.

$$TDS = [(Asymmetry * 1.3) + (Border * 0.1) \\ + (Color * 0.5) \\ + (Differential\ Structures * 0.1)] \quad (1)$$

Tumors being given a TDS higher than 5.45 are considered highly suggestive of melanoma, an excision is recommended for tumors with a TDS higher than 4.8.

In order to build dermatologists' models of diagnosis/therapeutic decision, five senior dermatologists were asked to give their diagnosis and therapeutic decision for 227 images of tumors, together with their opinion about the existence of malignancy features (presence/absence). 'Models' of dermatologists were subsequently built by connecting predicted features to the so-called "gold standard" diagnostic (see below).

2   Materials and methods

The initial dataset used in this study was collected at the dermatology departments of the British Hertfort Hospital and the Louis Mourier Hospital in 'Ile de France' (France). A total of 900 images of pigmented skin lesions were

acquired in 'uncontrolled' conditions (see [7]). As a consequence of the inclusion protocol, many tumors were quite similar, and melanomas were largely in a minority. The current working database that initially included all identified melanoma lesions has been completed to 227 with randomly selected tumors. On doing so, it appeared that 77 lesions were classified as benign lesions. In order not to cause any needless distress to the patient, the majority of benign lesions were not surgically excised. Dysplastic lesions (i.e. atypical lesions, for which malignity may be suspected) were 118 in the database. Thirty-two pigmented lesions were categorized as malignant melanomas. The malignant melanomas and the dysplastic lesions were all surgically excised and histopathologically analyzed.

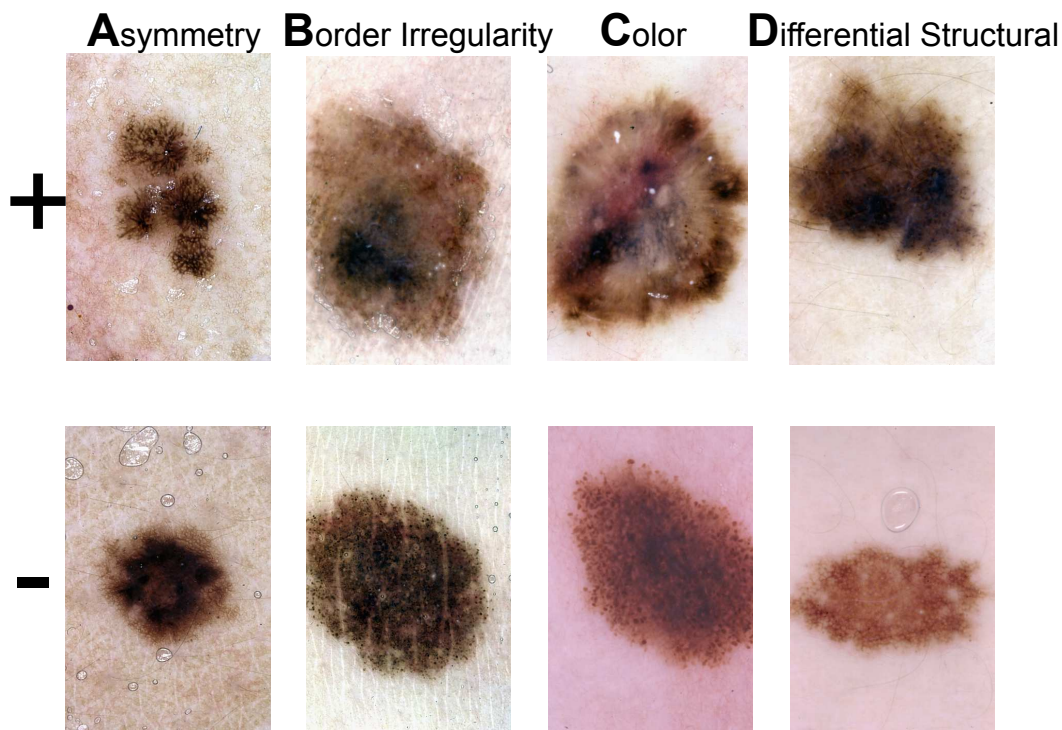### Asymmetry   Border Irregularity   Color   Differential Structural

Fig. 1. Four nevi that fulfill ABCD rule (+) and four others that do not (-).

For this study, two classes were finally considered: histologically confirmed melanomas on the one hand and the remaining lesions on the other. For simplicity, this classification is referred to as the 'gold standard' diagnosis in this study.

Five senior dermatologists were asked for their expertise about the 227 selected images. They were presented each tumor both as macroscopic image and dermoscopic image. They subsequently gave their opinion about the presence of ABCD and dermoscopic features (dichotomic answers), their diagnosis (melanoma, dysplastic or benign lesion) and their therapeutic decision (dichotomic answer, excision/non-excision). Mimicking the Stolz's linear decision model, a logistic regression classifier [8] was built for each dermatologist using the features they reported as input and the 'gold standard' diagnosis as output, while a leave-one-out cross-validation was employed. The classifiers provide a probability to be a melanoma for each tested lesion in the selected database. ROC curves were built from these probabilities. They allows further analyzing the whole set of sensitivity/specificity couples of parameters. The area under the ROC curves (AUC) is a measure of the quality of prediction.

3   Results

As far as the diagnosis is concerned, one may observe a high variability of sensitivity among dermatologists whereas specificity remains similar, with the exception of the one obtained by dermatologist 3 (Table I).

TABLE I
Dermatologists' performances

Diagnosis and therapeutic decision	Diagnosis Sensitivity/Specificity	Therapeutic decision Sensitivity/Specificity
Dermatologist 1	0.62 / 0.90	0.84 / 0.63
Dermatologist 2	0.78 / 0.85	0.93 / 0.63
Dermatologist 3	0.59 / 0.71	0.84 / 0.39
Dermatologist 4	0.81 / 0.90	0.84 / 0.55
Dermatologist 5	0.71 / 0.80	0.87 / 0.63

Sensitivity and specificity are calculated with respect to the 'gold standard' diagnosis of melanoma.

In fact, the analysis of dermatologists' performances requires considering several factors. Sensitivity and specificity express the efficiency of the clinicians, but also the trade-off they believe to be acceptable with respect to the risk for a false diagnosis (Table I).

Depending on their level of confidence, they may privilege sensitivity over specificity. The opposite may also be true since it is a "risk-free" trial. The prior frequency of melanoma they meet usually in their daily practice may also play a role.

All these factors also take part in the therapeutic decision, although a much smaller one. In fact, we can expect (and observe) the therapeutic decision to have a higher sensitivity (as far as the prediction of melanoma is concerned), together with a lower specificity, since the CDL worthy of an excision encompasses melanoma. At the therapeutic decision level, sensitivities are more comparable, most melanomas are detected, but the cost (specificity) highly varies from one dermatologist to another.
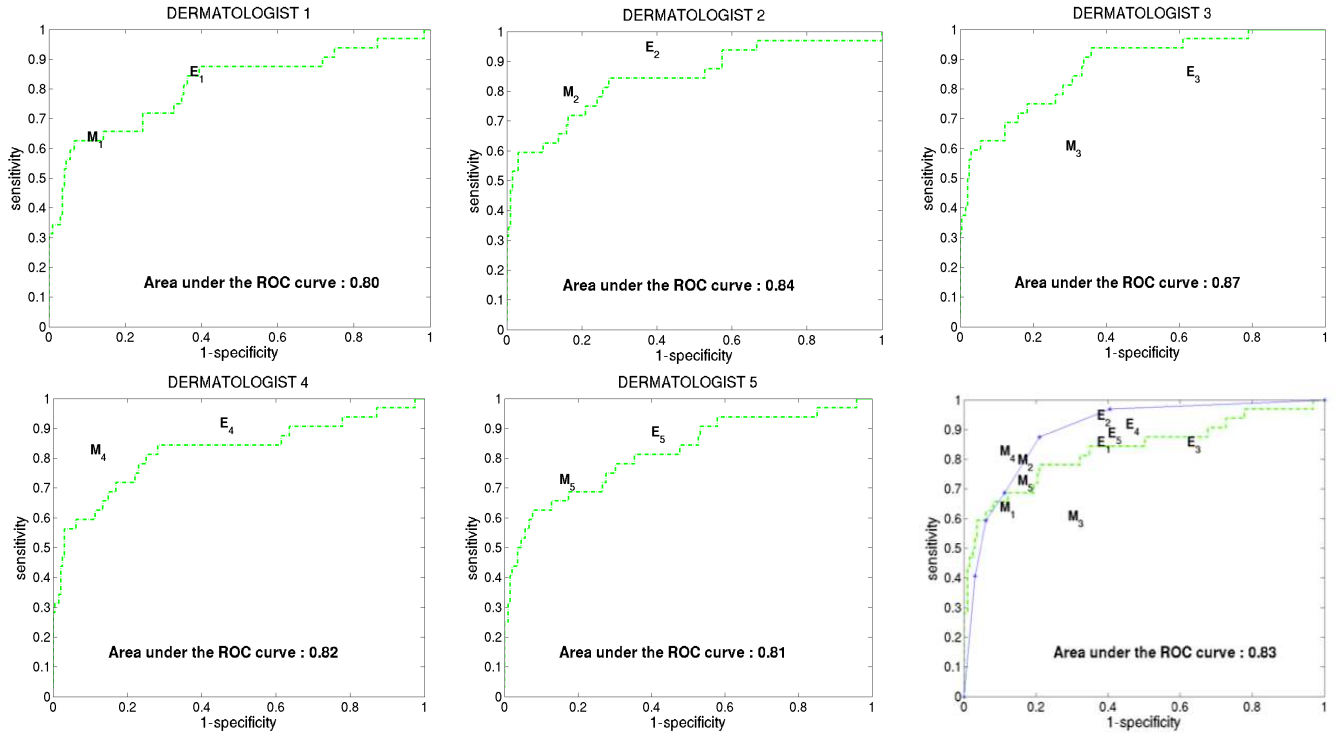

Fig. 2. Roc Curves for melanoma diagnosis result from logistic regression based on the features detected by each dermatologist. M_i and E_i show the accuracy of dermatologist i' diagnosis and therapeutic decision (Panels 1 to 5).
The last panel (bottom right) shows the Roc Curve of the logistic regression based on the consensual detected features (dotted line), together with the diagnosis and the therapeutic decision of each dermatologist. The 5-point solid line results from the voting schema about diagnosis so that the lower point corresponds to the tumors reported as melanoma by each of the 5 dermatologists, the next point corresponds to the tumors reported as melanoma by 4 out of the 5 dermatologists and so on.

Dermatologist' performances are shown, one at a time, in the subplots of Fig. 2. Sensitivity and specificity are displayed together with a ROC curve obtained with the mentioned linear classifier. It can be seen that dermatologist 1 grounds its diagnosis on the mere basis of the features he detected. Dermatologists 2, 4 and 5 probably use of additional visual features not available to the classifier, which makes their diagnosis and therapeutic decisions better than the results obtained by the classifier. Finally, dermatologist 3 seems poorly combining the features he has however efficiently detected. The best classifier performance is obtained from the set of features detected by the dermatologist 3, as shown by the AUC, which is the highest in this study.

Combining dermatologists' diagnoses and features characterization allows evaluation of the efficiency of the group of experts together. As dermatologists do not necessarily agree about the presence of features, diagnosis and therapeutic decision, a voting schema has been implemented (see reference [7] for details). It showed that full agreement between dermatologists is high (60%) as far as diagnosis is concerned, whereas therapeutic decision is more disputed (36%) (Table II). The picture is contrasted for the features: The agreement is high for asymmetry and relatively poor for color irregularity (Table II).

TABLE II
Distribution of the 227 images for ABCD features as a function of the dermatologists' vote

Feature	0-5	1-4	2-3
Asymmetry	54%	26%	20%
Border	36%	38%	26%
Color	34%	38%	28%
Differential structure	46%	30%	24%
Diagnosis	60%	23%	17%
Excision	36%	36%	28%

0-5 indicates that the 5 dermatologists are in full agreement, 1-4 indicates that 1 out of the 5 dermatologists disagrees and so on.

Combining diagnosis provides a remarkable result (Fig. 2, last panel, highest point of the 5-point solid line): 31 out of 32 melanomas are detected (sensitivity = 0.97) while cost remains low (specificity = 0.60). In contrast, the "consensual" ROC curve (AUC = 0.83) provided by the logistic model based on the consensual detected features does not reach the best available performance (0.87, Fig 2). Finally, the Stotz's formula, lightly adapted to fit our protocol, get an AUC of 0.79, which is quite good in this context.

4 Conclusion

In this study, five senior dermatologists were asked for their expertise about the 227 selected images. Models of diagnosis and therapeutic decision based on the observations of the presence of ABCD and dermoscopic features have been presented and evaluated. The results obtained show that the variability of performance of dermatologists is high, dermatologists with a melanoma-specific hospital activity showing the best performance, both for the diagnosis and the therapeutic decision.

The sensitivity and the specificity for diagnosis as well as therapeutic decision are higher if clinicians' advices are pooled. Such a result was not always assured, given the false positives to be cumulated.

Models also allow observing to what extent dermatologists ground their diagnosis on the malignancy features they detected. We believe that the clinical experience (based on the learning by sample paradigm) they gain during their daily practice is the key to their success.

5 Acknowledgment

6 References

[1] H. Kittler, H. Pehamberger, K. Wolff, M. Binder, "Diagnostic accuracy of dermoscopy", *Lancet Oncol*, Vol. 3, pp. 159– 165, 2002.

[2] F. Nachbar, W. Stolz, T. Merkle, A. Cognetta, T. Vogt, M. Landthaler, P. Bilek, O. Braun-Falco, G. Plewig, "The ABCD rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions", *Journal of the American Academy of Dermatology*, Vol. 30(4), pp. 551–559, 1994.

[3] R. Feldmann, C. Fellenz, F. Gschnait. "The ABCD rule in dermatoscopy: analysis of 500 melanocytic lesions", *Hautarzt*, Vol. 49, pp. 473–76, 1998.

[4] W. Stolz, A. Riemann, B. Armand, et al. "ABCD rule of dermatoscopy: a new practical method for early recognition of melanoma", *Eur J Dermatol*, Vol. 4, pp. 521–27, 1994.

[5] G. Capdehourat, A. Corez, A. Bazzano, P. Musé, "Pigmented skin lesions classification using dermatoscopic images", *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, Vol. 5856, pp. 537-544, 2009.

[6] R.H. Johr, "Dermoscopy: alternative melanocytic algorithms? The ABCD rule of dermatoscopy, menzies scoring method, and 7-point checklist", *Clinics in Dermatology*, Vol. 20(3), pp. 240–247, 2002.

[7] A. Tenenhaus, A. Nkengne, J.F. Horn, C. Serruys, A. Giron, B. Fertil, Detection of melanoma from dermoscopic images of naevi acquired under uncontrolled conditions, *Skin Research and Technology*, Vol. 16(1), pp. 85-97, 2010.

[8] J.G. Liao, Khew-Voon Chin. Logistic regression for disease classification using microarray data: model selection in a large p and small n case, *Bioinformatics*, Vol. 23 (15), pp. 1945-1951, 2007.