Arch Dermatol Res (2010) 302:551–560
DOI 10.1007/s00403-010-1051-6

ORIGINAL PAPER

Objective follow-up of atypical melanocytic skin lesions: a retrospective study

Pietro Rubegni · Gabriele Cevenini · Marco Burroni · Riccardo Bono · Paolo Sbano · Maurizio Biagioli · Massimiliano Risulo · Niccolò Nami · Roberto Perotti · Clelia Miracco · Michele Fimiani

Received: 13 November 2009 / Revised: 29 March 2010 / Accepted: 6 April 2010 / Published online: 22 April 2010
© Springer-Verlag 2010

Abstract Various authors have suggested that information from longitudinal observation (follow-up) of dynamic changes in atypical melanocytic pigmented skin lesions (MPSL) could enable identification of early malignant melanoma escaping initial observation due to an absence of specific clinical and dermoscopic features. The aim of our retrospective study was to determine the existence of numerical variables regarding changes in MPSL that could be useful to differentiate early melanomas and atypical nevi. The study was carried out in two Italian dermatology Centres. Digital dermoscopy analyzers (DB-Mips System) were used to evaluate dermoscopic images of 94 equivocal pigmented skin lesions under observation for 6–12 months and then excised because of changes across time (29 melanomas and 65 nevi). The analyzer evaluates 49 parameters grouped into four categories: geometries, colours, textures and islands of colour. The ROC curve designed on the 49 digital dermoscopy analysis parameters showed good accuracy. At sensitivity (SE) = specificity (SP), it correctly classified 89.3% of cases. When objective pigmented skin lesion parameters were considered together with their objective changes over 6–12 months, a decisive increase in discrimination capacity was obtained. At SE = SP accuracy was 96.3%. Analysis of the parameters of our model and statistical analysis enabled us to interpret/identify the most significant factors of modification and differentiation of lesions, providing quantitative insights into the diagnosis of equivocal MPSL and demonstrating the utility of objective/numerical follow-up.

Keywords Melanoma · Pigmented skin lesions · Dermoscopy · Digital dermoscopy · Multivariate analysis

P. Rubegni (✉) · M. Burroni · P. Sbano · M. Biagioli · M. Risulo · N. Nami · R. Perotti · M. Fimiani
Dermatology Section,
Department of Clinical Medicine and Immunological Science,
Policlinico "Le Scotte", Viale Bracci, 53100 Siena, Italy
e-mail: rubegni@unisi.it

G. Cevenini
Department of Surgery and Bioengineering,
University of Siena, Siena, Italy

R. Bono
Istituto Dermopatico dell'Immacolata, Rome, Italy

C. Miracco
Department of Human Pathology and Oncology,
University of Siena, Siena, Italy

Introduction

The best way to improve the prognosis of melanoma is believed to be early diagnosis, as thinner melanomas are associated with substantially less mortality than thicker tumours [15, 53]. However, in its initial stages differential diagnosis of malignant melanoma (MM) and benign clinically atypical melanocytic pigmented skin lesions (MPSL) is particularly challenging (SPSS Advanced Models™ 10.0) [9, 33, 52]. Dermoscopy, also known as epiluminescence light microscopy (ELM), a non-invasive method for evaluation of MPSL, proved able to improve diagnostic accuracy for del MM [50, 51]. Two systematic reviews have demonstrated that ELM increases diagnostic accuracy for MM by up to 35–49% with respect to examination with the naked eye [36]. Although ELM has greatly improved the possibility of differentiating benign MPSL from MM, it is still very difficult to distinguish very atypical nevi and

early MM with certainty. This is not surprising, because this aspect has not even been solved histologically. As well, 10–20% of melanoma lack the criteria that normally suggest the diagnosis during clinical/dermoscopic examination. This means that about one-fifth of cases of melanoma cannot be diagnosed by the present methods [5, 10, 14, 40, 41, 43, 49]. The only clue to diagnosis of these lesions is often the history of morphological changes reported by patients [4, 18, 42, 44]. Various authors have suggested that additional information from longitudinal observation (follow-up) of dynamic changes in atypical MPSL could enable identification of MM escaping initial observation due to an absence of specific clinical and/or dermoscopic features [2, 29, 30, 48]. Recent widespread use of digital photography and software for storing images has greatly facilitated follow-up also from a technical point of view [11]. However, concern has been expressed that digital follow-up of suspicious or equivocal MPSL, rather than immediate excision, could delay the treatment of melanoma [13, 27, 47]. This, however, is only true for lesions that at first observation had clinical/dermoscopic features at least sufficient to suspect the possibility of MM. Besides, it is unthinkable to excise all lesions which are atypical but do not fulfil the criteria for MM diagnosis. On the other hand, digital-ELM follow-up of lesions having equivocal clinical/dermoscopic features has been shown to identify early MM that had not yet acquired ELM features typical of MM [23]. This seems to be confirmed by a recent paper by Haenssle et al. [22] who reported that long-term follow-up of equivocal MPSL by digital-ELM increased detection sensitivity (SE) for MM.

The recent introduction of digital-ELM and sophisticated image processing software [digital dermoscopy analysis (DDA)] has opened a new dimension in the evaluation of benign and malignant MPSL, allowing visualization and recognition of slight morphological and colorimetric differences that are difficult to quantify with the naked eye [2, 21]. Since this method is based on numerical description of clinical images, it allows unequivocal comparison of MPSL images in time (follow-up). Digital dermoscopic monitoring of melanocytic lesions could help to minimize excision of benign lesions.

In the present study, we retrospectively reviewed DDA on a series of equivocal MPSLs that had been followed up for variable periods and then excised on the basis of specialist opinion. The aim was to determine the existence of numerical variables regarding changes in MPSLs that could be indicative of malignant evolution. Stepwise Bayesian discriminant analysis was used to select an optimal minimum subset of statistically significant variables for model classification purposes. The predictive power of the most discriminating models was evaluated with respect to histological diagnosis.

Materials and methods

All clinical investigation was conducted according to Declaration of Helsinki principles. Institutional approval and patient consent were obtained for all experimental procedures. This retrospective study was carried out in two Italian Centres: the Department of Dermatology at Siena University and the Istituto Dermopatico dell'Immacolata in Rome between September 2007 and December 2008. Ninety-three clinically equivocal MPSLs, followed up for 6–12 months before excision (mean follow-up 8 months), were the subject of the study (male female ratio of patients 1.1:1; mean age 39 years, age range 28–57 years) (Fig. 1). Twenty-nine out of 93 were excised, flat or superficially raised MM followed up for at least 1 year (18 in situ MM: thickness ranging from 0 to 1.6 mm and median invasion depth of invasive melanomas 0.4 mm) and 64 were excised, dermoscopically atypical (irregular/prominent network and irregular overall pigmentation) benign PSL (43 dysplastic nevi, i.e., nevi with architectural disorder and cytological atypia. Three/four benign lesions were matched for period of follow-up, patient age, sex, phototype, number of nevi (more or less than 50) and presence/absence of freckles with each MM. Genital, palmo-plantar and head lesions were not included in the study because they have different particular dermoscopic features. Histories of trauma and UV lamps were other exclusion criteria. At the first examination, MPSLs were atypical but did not meet dermoscopic MM criteria (ABCD dermoscopic algorithm) [11]. All these lesions were removed during follow-up because they changed with time, even slight changes considered clinically and dermoscopically significant, identified by classical digital dermoscopy (16–40×) performed by expert clinicians (P. R. and R. B.). Histopathological diagnosis was based on NIH Consensus Conference criteria [37]. Slides of the lesions were examined by three dermopathologists (C. M., P. R. and M. F.). They were classified as melanoma or nevi when at least two out of three dermopathologists agreed on the diagnosis [37, 45–47, 54]. We observed a percentage discordance of about 20%.

Measurements

The lesions were imaged (magnification 16×), stored and analysed by the DB-Mips System (Biomips Engineering, S.R.L., Siena, Italy), a computerized instrument providing a visual database and objective evaluations of pigmented skin lesions. In our medical records we have about 57,000 images of skin lesions from 12,000 patients. Before beginning the study, the system was tested on 87 common nevi for which we had two images taken 6–12 months apart.

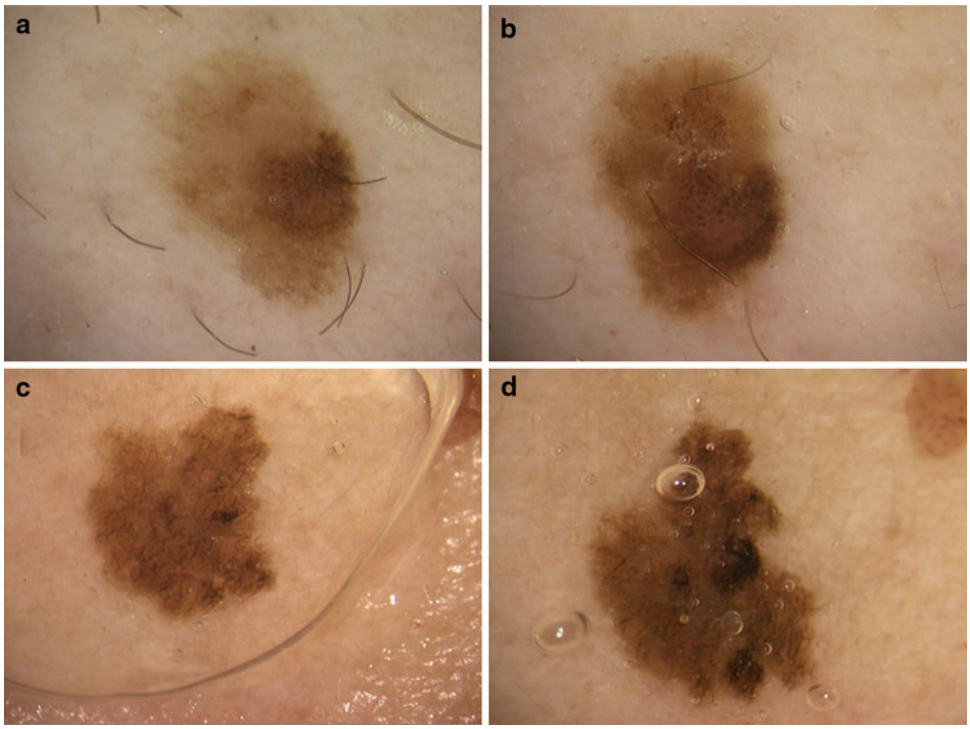

Fig. 1 Two illustrative cases of melanoma of our study population followed for 6 (**a** before and **b** after) and 12 months (**c** before and **d** after)

Variations exceeding 5% between before and after were not observed in any studied variable. The patient lay on the examination table with the skin surface around the lesion orthogonal to the incident light. If the lesion and/or surrounding skin was hairy, the hairs were carefully removed with scissors or a razor. The lesion was recorded as a digital signal and saved. After at least 6–12 months of follow-up, the clinicians (PR and MR) decided to excise the lesions as suspected MM. The lesions were removed surgically (by PS, MB and PR) and histological examination was performed. All digital images were analysed using appropriate algorithms.

Equipment

The DB-Mips System consists of a 3CCD (charge coupled device) PAL Broadcast video camera with 730 lines of image resolution and 60 dB signal-to-noise ratio. The camera was connected to a handheld 3CCD optical system with five magnifications from 6 to 40, enabling fields from 4 mm to 4 cm in size. The camera was calibrated weekly using special paper for white balance. Illumination was provided by a 150-W light source at 3,200 K. The components of the video signal were connected to a frame-grabber interfaced with a Pentium III 500 MHz Personal Computer having a magneto-optical drive for image storage. The system ran under Microsoft Windows, and all the software was written in language C/C++.

Digitization and parametrization

Choice of the most useful features to extract from digital images depends on the results of epiluminescence pattern analysis. The variables we chose were dermoscopic parameters currently used in the diagnosis of MPSL. Although the system saves microscope magnifications along with texture analysis, offering an objective evaluation, the different magnifications could confuse clinicians wanting to make subjective comparisons of lesions. In this paper we therefore only discuss images with a magnification of 16×. The system used a procedure for digital image processing based on the Laplacian filter for segmentation and a zero-crossing algorithm for border automatic outline. It then evaluated 49 parameters (derived from classical dermoscopy) [46, 47] for discriminating power. Reproducibility was first tested on digitized images of 50 lesions belonging to 34 subjects (1 MPSL per patient recorded 5 times at 15 min intervals). Absolute differences between single measurements and mean values of a given lesion or parameter never exceeded 5% of the mean value. The parameters, as previously described, belonged to four categories: geometries, colours, textures and islands of colour (i.e., colour clusters inside the lesion) [47].

Statistical analysis

Multivariate analysis of variance (MANOVA) for repeated measures was used to evaluate the statistical significance of

changes in the digital dermoscopy variables of suspected malignant skin lesions. Measurements were repeated 6 and/or 12 months after (time T2) an initial reference observation (time T1) [6, 31]. Two types of lesions were compared: excised melanomas and excised atypical nevi. MANOVA included examination of *within-subject* and *between-subject* effects, univariate analysis of each numerical (digital) dermoscopic variable and interactions between lesion types and their changes [6]. Statistical significance was set at 95% ($P < 0.05$). MANOVA statistical computations were performed using SPSS software Advanced Models™ 10.0.

Diagnostic model design

The diagnostic ability of digital dermoscopy to discriminate between MM and benign atypical MPSL was estimated by multivariate linear Bayesian classification analysis [7, 8, 19, 49]. Two discrimination models were designed: they were learned on numerical dermoscopic variables measured at, T2 (model 1) and both times (model 2). Model 2 was trained on parameters evaluated at time T2 and differences between measurements at T2 and T1. An essential subset of variables with the highest statistically significant discrimination power was identified using a stepwise procedure guided specially to select only clinically relevant features [18, 25–27, 33, 36, 39, 40]. We applied the leave-one-out (LOO) cross-validation procedure, also known as jackknife method, to ensure suitable model generalization in the correct classification of benign and malignant skin lesions. LOO uses all available data to train and test models: it classifies each lesion in turn using the discrimination model constructed with all other cases and has been proven to give a reliable estimation of prediction error [17, 19, 49]. Feature selection was carried out by maximizing the area under the receiver operating characteristics (ROC) curve (AUC), calculated on LOO testing data and stopping the stepwise process when AUC did not show any further statistically significant increase [24, 32, 38]. The accuracy of the Bayesian model was also evaluated at the ROC curve point of equal SE and specificity (SP) giving the percentage of correctly classified lesions CCL%. The discrimination power of the three models was compared through ROC curves and their 95% confidence intervals which were estimated by resampling from available data through the Monte Carlo computational method [39]. Resampling was carried out by numerically simulating data extracted from a multinormal distribution fitted to nevus and melanoma training samples. One thousand samples, with the same number of cases as the experimental data, were simulated. The population ROC curve was obtained averaging the 1,000 simulated sample ROC curves. 95% CI was obtained taking the 2.5th and 97.5th percentiles of 1,000 curves. A convenient smoothing effect on CI curves was introduced by computing the 2.5th percentile as the mean of percentiles 0–5 and the 97.5th percentile as the mean of percentiles 95–100. Estimated model ROC curves were analysed to consider convenient clinical decision strategies. Bayesian model design and Monte Carlo resampling technique were performed using Matlab software [35].

Results

MANOVA analysis is reported in Tables 1 and 2, which show statistical comparisons between lesion *types* and between observation *times*, respectively. Table 1 shows significant differences in mean lesion dimensions (area, perimeter and diameters) between nevi and melanomas. Statistical interactions between lesion types and annual changes are reported in the right-hand part of Table 2. For area, percentage changes between T1 and T2 (6–12 months percentage change, *C%*) were also significant (*C%* was about zero for nevi and 25.3% for melanomas). Less but still significant between-lesion differences were observed for perimeter, minimum diameter and maximum diameter. Contrast and entropy showed no between-lesion differences at either time but revealed significantly greater time changes in melanoma (18% for contrast and 6.8% for entropy). Thus, these parameters do not differ statistically between type of lesions at either time, but differ in *C%*, though of limited statistical significance ($F = 5.0$); this could indicate that melanoma tends to differentiate from atypical nevi in time by developing greater contrast and entropy. Fractality of contrast, dark area, peripheral dark area, imbalance, light red area, number of red percentiles, number of blue percentiles, variance of border gradient and variance of border interruptions showed significant differences between nevi and melanomas at both times (number of red percentiles only at T2). However, these parameters did not show changes (*C%*) between T1 and T2, statistically different between MM and benign MPSL. Parameters quantifying multicomponent pattern, i.e., RED, green and blue multicomponent, showed a similar pattern to the above geometric parameters: melanomas systematically gave significantly higher parameter values and 6–12-month percentage changes than atypical nevi. Specifically, their sample values were about 40% higher than nevi at time T1 and about 55–90% at time T2, with 20–40% greater 6–12-month changes.

Table 2 illustrates the statistical significance of separate parameter differences for atypical nevi and melanomas between times T1 and T2. The geometric parameters area, minimum diameter and maximum diameter only increased significantly in melanomas. The red and blue component of nevus colour also increased (6–12 months percentage change *C%*) by 5 and 9.8%, respectively. Since in the RGB

Table 1 Multivariate analysis of variance (MANOVA): *between-lesion* analysis of dermoscopic parameters

Parameters	Time T1			Time T2			Time–lesion interactions			
	F	*P*	*D%*	*F*	*P*	*D%*	*F*	*P*	Nevi *C%*	Melanomas *C%*
Area	19.0	S	62	49.7	S	156	17.1	S	1.3	25.3
Perimeter	26.2	S	72.0	51.1	S	88.0	2.3	NS	–	–
Minimum diameter	13.4	S	39.6	37.9	S	59.8	14.5	S	0.9	15.5
Maximum diameter	20.2	S	46.6	50.2	S	69.3	20.6	S	−0.3	15.2
Contrast	0.2	NS	–	3.6	NS	–	5.0	S	2.5	18.0
Entropy	0.6	NS	–	1.9	NS	–	5.0	S	1.4	6.8
Fractality of contrast	5.0	S	−3.4	8.9	S	−4.5	0.7	NS	–	–
Dark area	6.4	S	51.8	9.1	S	60.3	0.4	NS	–	–
Peripheral dark area	7.8	S	62.0	10.7	S	75.3	0.5	NS	–	–
Imbalance	5.5	S	38.5	10.6	S	51.9	0.3	NS	–	–
Light red area	13.3	S	−396	12.8	S	−69.8	0.3	NS	–	–
Number of red percentiles	1.8	NS	–	6.3	S	29.7	0.8	NS	–	–
Number of blue percentiles	9.3	S	58.5	16.7	S	77.0	1.4	NS	–	–
Red multicomponent	12.6	S	41.8	56.5	S	89.8	41.1	S	0.4	34.3
Green multicomponent	8.7	S	38.6	34.9	S	74.5	16.5	S	−1.4	24.2
Blue multicomponent	3.5	NS	–	13.7	S	55.1	5.1	S	1.6	22.0
Variance of border gradient	4.9	S	38.3	7.3	S	48.4	0.1	NS	–	–
Variance of border interrupt.	7.4	S	−10.7	7.2	S	−10.0	0.0	NS	–	–

Differences were evaluated at times T1 and T2, and with respect to 6–12 months parameter changes. 6–12 months parameter changes as percentage of their initial values (time T1), *C%*; parameter differences between melanomas and atypical nevi as percentage of nevi value, *D%*. Only statistically significant parameter changes are reported

F Fisher value, *S* statistically significant ($P \leq 0.05$), *NS* not statistically significant ($P > 0.05$)

Table 2 Multivariate analysis of variance (MANOVA): *within-lesion* analysis of dermoscopic parameters

Parameters	Nevi			Melanomas		
	F	*P*	*C%*	*F*	*P*	*C%*
Area	0.2	NS	–	6.4	S	25.3
Minimum diameter	0.2	NS	–	8.6	S	15.5
Maximum diameter	0.0	NS	–	10.0	S	15.2
Red average	6.3	S	5.0	0.0	NS	–
Blue average	7.7	S	9.8	2.7	NS	–
Contrast	3.6	NS	2.5	12.1	S	18
Entropy	2.0	NS	–	6.0	S	6.8
Light red area	0.1	NS	–	5.3	S	100
Red multicomponent	0.0	NS	–	21.8	S	34.3
Green multicomponent	0.1	NS	–	10.6	S	24.2
Blue multicomponent	0.1	NS	–	4.5	S	22.0

6–12 months differences were evaluated separately for nevi and melanomas. 6–12 months parameter changes as percentage of their initial values (time T0), *C%*. Only statistically significant parameter changes are reported

F Fisher value, *S* statistically significant ($P \leq 0.05$), *NS* not statistically significant ($P > 0.05$)

system red is correlated with darkness/lightness we can assert that nevi interior pigmentation showed an increasing trend. Entropy only increased significantly in melanomas ($C\% = 6.8\%$). Light red area increased significantly in both lesions: $C\% = 33.4\%$ in nevi, 100% in melanomas. Parameters correlated with multicomponent pattern (red, blue and green multicomponent) only changed significantly in melanomas: red multicomponent $C\% = 34.4\%$, green multicomponent $C\% = 24.2\%$ and BLUE multicomponent $C\% = 22\%$.

Bayes diagnostic models

The diagnostic power of DDA parameters and their related 6–12-month follow-up is shown in Figs. 2 and 3 where the ROC curves of the three Bayes discriminant models are depicted. The continuous line represents the population ROC curve estimated by the Monte Carlo technique of multinormal data resampling; dotted lines show the 95% CI of the ROC curve and the dashed line indicates the experimental sample ROC curves. The intersections between the ROC curves and the diagonal line identify points of equal SE and SP.

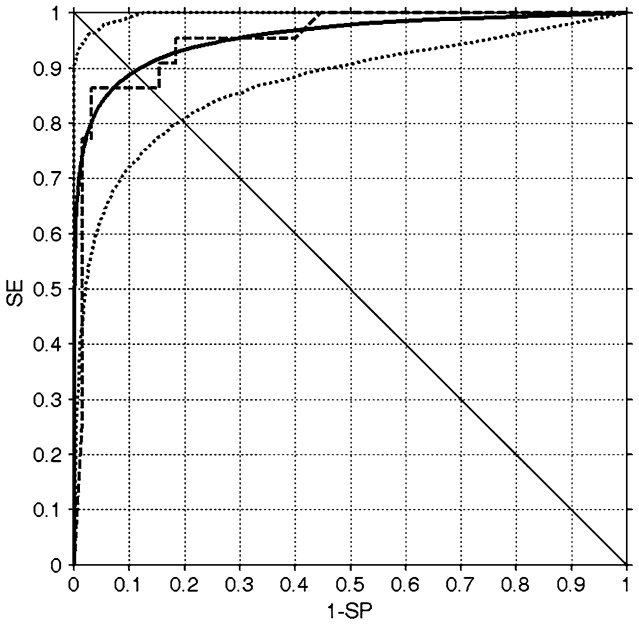

Fig. 2 Receiver operating characteristic (ROC) curve for Bayesian linear model 2 (dermoscopic parameters measured at time T2). Population ROC curve (*continuous line*) and its corresponding 95% confidence interval (*dotted lines*) estimated by Monte Carlo resampling numerical technique. *Dashed line* represents ROC curve obtained from experimental training data. *Diagonal line* represents points of equal sensitivity (SE) and specificity (SP); area under ROC curve (AUC); confidence interval (CI); estimated population AUC (95% CI) = 0.958 (0.912–0.988). Point of equal SE and SP = 0.893

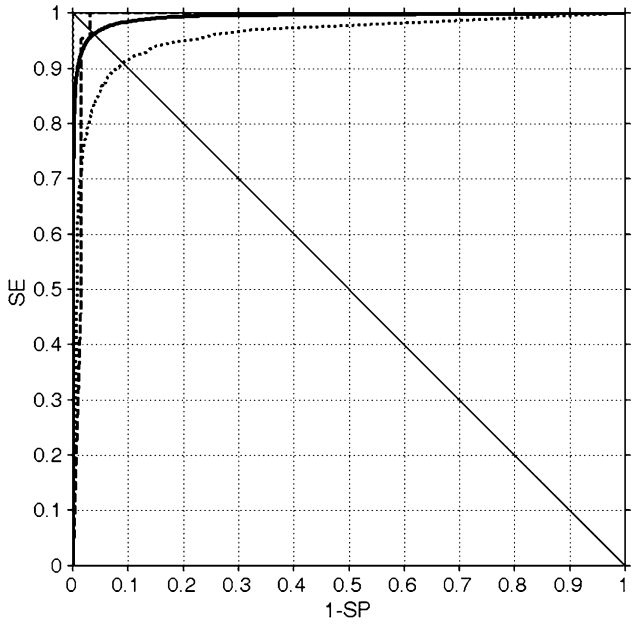

Fig. 3 Receiver operating characteristic (ROC) curve for Bayesian linear model 3 (dermoscopic parameters measured at time T2 and their differences from corresponding measurements at time T1). Population ROC curve (*continuous line*) and its corresponding 95% confidence interval (*dotted lines*) estimated from Monte Carlo resampling numerical technique. *Dashed line* represents ROC curve obtained from experimental training data. *Diagonal line* represents points of equal sensitivity (SE) and specificity (SP); area under ROC curve (AUC); confidence interval (CI); estimated population AUC (95% CI) = 0.995 (0.976–1.000). Point of equal SE and SP = 0.963

Bayes model 1, designed on dermoscopic data measured at T2 gave a population AUC of 0.958 and 95% CI of 0.912–0.988. The point at SE = SP on the ROC curve of Fig. 2 shows clearly that dermoscopic parameters measured at T2 lead to a population-estimated percentage of correctly classified lesions, CCL%, equal to 89.3. The lower bound of CI gave a CCL% of about 80 or 20% of misclassified lesions. In other words, on the basis of our experimental data, misclassified lesions can be as high as 20%.

When dermoscopic parameters measured at T2 were considered together with their changes over the previous 6–12 months (Bayes model 2), a decisive increase in discrimination capacity was obtained. The population ROC curve of model 3 and its 95% CI are shown in Fig. 3. Accuracy was high: estimated population AUC was 0.995 and 95% CI ranged from 0.976 to 1. The CCL% at SE = SP was between 92% (lower bound of 95% CI) and 100%. This means that in the worst case only 8% of lesions would be misclassified. To minimize the number of false-negative lesions, it is customary to identify a point on the ROC curve that corresponds to an SE > SP. For example, where SP = 70% (30% of false positives), the estimated population SE was very close to 100% and the lower bound of 95% CI was still about 97%.

The right part of Table 3 reports the stepwise procedure for designing model 2. An optimal set of six predictor parameters was identified: area and blue average measured at T2 and annual changes in red multicomponent, blue average, ellipsoidality and variance of border interruptions. Obviously, when considered separately in the multivariate analysis, these variables lose their intrinsic meaning. They are nevertheless useful for constructing statistical classifiers.

Discussion

In treating melanoma, the clinical dermatologist's principal aim is to remove all malignant MPSL as early as possible, while avoiding unnecessary biopsies [15]. This aim is complex for two main reasons: (1) no method (naked eye, ELM, DDA, reflectance confocal microscopy) has a SE of 100%; (2) many atypical MPSL have clinical features resembling early melanoma [1, 3, 5, 11, 20, 23, 34]. According to many authors, the only possible course of action is photographic comparison, now considered the best clinical approach for detecting developing melanoma [2, 20–22]. Puig et al. [41] recently reported that as many as 67% of a series of 97 MM

Table 3 Design of Bayes linear models from experimental data: stepwise selection of dermoscopic parameters

Step no.	Model 1: time T2			Model 2: T2 plus T1–T2 changes		
	Parameters	AUC	CCL%	Parameters	AUC	CCL%
1	Area	0.891	81.7	Area (time T2)	0.891	81.7
2	Blue average	0.917	83.5	Red multicomp. (time changes)	0.941	87.9
3	Green average	0.952	89.0	Blue average (time changes)	0.962	89.7
4	No. of green percentiles	0.958	89.3	Ellipsoidality (time changes)	0.966	90.8
5				Dark area (time T2)	0.978	92.2
6				Variance of border interrupt. (time changes)	0.995	96.3

Estimated population area under ROC curve, AUC and percentage of correctly classified lesions at ROC point of equal sensitivity and specificity, CCL%

difficult to diagnose were removed on the basis of a history of changes noted by the patient or by the clinician during dermoscopic follow-up. We therefore decided to objectively assess the degree to which follow-up can contribute to differentiation of equivocal malignant and benign MPSL in terms of diagnostic accuracy. To test whether discrimination models maintained the same predictive performance on all data, including data not used to construct the model, we used the cross-validation LOO technique. This method is particularly useful in biomedical classification when little data is available. Sacrificing a part of the data to test model predictive ability may compromise the validity of a model, because the model is trained on too few cases. LOO efficiently uses all data for training and testing model predictive performance at little higher computational expense. It trains as many models as the data available, classifying the case left out each time with the model trained on the remaining cases.

Objective comparison of nevi and melanomas at T2 (day of exeresis)

Our results showed that it was possible to detect statistically significant differences in some of the variables between benign atypical MPSL and equivocal MM at T2, i.e., the moment when the clinician decided to remove the lesions. Significant differences in the mean lesion dimensions (area, perimeter and diameters) were found between nevi and MM. MMs were larger than nevi (Table 1), a unanimously accepted fact among the clinical criteria for differential diagnosis of MM [1]. Our study confirms that geometric variables are still very important in the differentiation of atypical nevi and early melanomas. Other significant variables concerned the disposition and imbalance of colour inside lesions (dark area, peripheral dark area, number of blue percentiles, imbalance). This is in line with the results of ELM studies showing areas of accumulation of dark pigment (black dots, brown globules, diffuse grey–black blotches) to be more frequent in MM than in benign atypical MPSL. Border-related variables (variance of border gradient, variance of border interruptions) and light red area (an attempt to render objective milky-red areas in MPSL) were also significant. Likewise, previous studies suggested that dermoscopic border features by ELM and milky-red areas were important features for differentiating malignant and benign MPSL [3]. MANOVA at T2 also showed statistically significant differences for variables that we constructed to objectify multicomponent patterns (red multicomponent and green multicomponent). These variables, that express the number, dimensions and colour differences between objects within MPSL, are therefore correlated with the structural asymmetry and "disorder" evident in MM with respect to benign atypical MPSL [3].

The diagnostic capacity of DDA to discriminate between MM and benign atypical MPSL at T2 was therefore estimated by multivariate linear Bayesian classification analysis (model 1) designed by a stepwise procedure for DDA parameters selection. The objective dermoscopic parameters selected by the stepwise procedure represented the optimum minimum multivariate subset of parameters giving the highest statistically significant discrimination power. At T2, at the point of equal SE and SP, the population ROC curve classified 89.3% of cases correctly. Considering also the wide confidence intervals (Fig. 2), in our opinion, these values demonstrate that a 10–15% of lesions are clinically difficult to distinguish by DDA.

Within-lesion comparison of nevi and melanoma after 6–12 months (T1 vs. T2)

The second aim was to determine whether variables of change reflecting natural progression could be extrapolated among MM and among nevi by DDA. We analysed the changes occurring over 6–12 months and endeavoured to identify variables that changed in a statistically significant manner. Differences between values at T1 and T2 detected by within-lesion MANOVA were expressed as percentage change *C%* (Table 2). Differences between *C%* of MM and

benign MPSL determined the significance of variation of individual variables in the course of 6–12 months (Table 1). Within-lesion analysis indicated that MM increased in dimension (area, minimum and maximum diameter), manifested greater disorganization of internal components (red multicomponent, green multicomponent, blue multicomponent, contrast and entropy) and increased in milky pink component, if any (light red area). MM developed greater contrast and entropy. If we imagine a grid superimposed on a MPSL, contrast expresses differences between light and dark squares of the grid and entropy describes how irregular the disposition of light and dark squares is. These two correlated parameters, used to describe lesion texture, provided an indication of the "order" of patterns within lesions. Parameters describing the number of different pattern components in the three bands of colour (red, green and blue multicomponent), used to quantify the dermoscopic picture defined as "multicomponent pattern", showed a trend similar to the above geometric parameters: melanomas systematically gave significantly higher parameter values and 6–12 months changes than atypical nevi, showing increases (*C%*) of 22–34%. These were the parameters with the highest percentage changes. Interestingly, the percentage change was not significant for benign MPSL. Only two variables with borderline significance were found for benign MPSL. The existence of objective differences in evolution distinguishing MM from benign MPSL is therefore confirmed. These differences mainly regard variables reflecting the disorganization/organization of the pigment network and multicomponent pattern. This discovery suggests that in cases in which dermoscopic diagnosis is difficult at the first examination, dermoscopic follow-up of the MPSL will certainly be useful for detecting significant objective changes in variables associated with evolution of MM and not of benign MPSL (especially multicomponent pattern variables).

Comparison of DDA discrimination capacity for melanoma and nevi, adding the variable "change"

The third aim of the study was to assess whether percentage change could help in the diagnosis of MM. Again, this was done by multivariate linear Bayesian classification analysis (model 2). When dermoscopic parameter values measured at T2 were considered together with their objective changes over 6–12 months, as for Bayes model 2, a decisive increase in discrimination capacity was obtained. Model 2 showed very high accuracy: estimated population AUC was 0.995 and 95% CI ranged from 0.976 to 1. ROC curve point of equal SE and SP indicate that in the worst case, only 8% of lesions would be misclassified. Taking into account the problematic nature of studied lesions, i.e., "difficult to diagnose melanomas" and atypical nevi, and the unavoidable rate of disagreement in the histological diagnosis, a maximum error of 8% is sufficiently low for a reliable clinical application. Moreover, the error can be conveniently modulated by choosing a point on the ROC curve (Fig. 2) that decreases the number of false negatives (higher SE) at the expense of false-positives (lower SP). This allows the rate of misclassification of melanomas as nevi to be reduced. For example, admitting SE = 99.9% we have a SP of 63%. Thus, a considerable number of excisions could also be avoided [54]. However, we have to underline that among the false negative-cases we observed a nodular melanoma 1.6 mm thick (Fig. 1b). We can suppose two interpretations of this matter. The first is that this kind of lesions can be very difficult to diagnose [27]. In nodular melanoma, many of the classic dermoscopic features of SSM are usually lacking, especially those dermoscopic structures that correspond to the flat parts of the SSM [50]. The second could be that our system was not trained enough to recognise these kind of lesions (only one nodular melanoma in our study population). On these bases we suggest not to evaluate nodular lesions by DDA.

The following major observations emerge from our experience and results. First of all, in our opinion, the lesions to examine by DDA must be selected and this may be done with the help of dermoscopy. Only expert clinicians can use DDA when they believe it is needed, namely for equivocal MPSL. No instrument has a clinician's ability to differentiate a traumatized seborrhoeic wart from a melanoma, or to diagnose amelanotic melanoma. Regarding time spans and intervals for monitoring small dubious lesions, like those studied by us, it is most unlikely that changes can be appreciated in <6 months, though this is obviously not a universal rule. Finally, regarding the time requirements and cost effectiveness of the method, the equipment (digital camera, dermoscopic adaptor, computer and software) cost 2000–3000 € and the time taken to acquire an image and for the software to evaluate it, using WiFi Technology, was about 15 s.

In conclusion, the statistical models designed to differentiate equivocal benign and malignant MPSL based on objective variables had no completely satisfactory accuracy, misclassifying about 10–20% of lesions, as found by other authors. On the other hand, using changes in variables over the course of 6–12 months, a model with embarrassingly high classification power (diagnostic accuracy ranging statistically from 92 to 100%) was obtained. Analysis of the parameters of this model and MANOVA enabled us to interpret/identify the most significant factors of modification and differentiation of lesions, providing quantitative insights into the diagnosis of equivocal MPSL and demonstrating the utility of objective/numerical follow-up [12, 16, 28, 45].

It is well known that changes in MPSLs over time may be important for differential diagnosis of nevi and mela-

noma. Indeed, the parameter "evolution" has been added to a common diagnostic algorithm for melanoma, ABCDE. However, there is no detailed objective data on which clinicians can rely to know exactly what changes (colour, dimensions, asymmetry, number or colours) are more indicative or typical of malignant behaviour or malignant trend and should therefore particularly be assessed. The present study provides indications of this type. The data obtained demonstrate that when atypical MPSLs are observed by DDA over a period of time, their behaviour may change but the entity and quality of changes are different for nevi and MM. According to our observations, the elements most useful for differentiating the behaviour of benign and malignant lesions in time are geometric variables expressing changes in lesion dimensions and certain parameters describing the type and distribution of colour inside the lesions. These differences may also be seen by simple clinical examination which can be made more effective by appropriate photographic follow-up. The suggested procedure offers at least two advantages: it avoids unnecessary excisions (nevi change little and in a typical manner), and it identifies at least some featureless MMs. However, the procedure has two limits: it cannot be extended to MPSLs in areas such as the scalp, face, palms, soles and mucous membranes, and it is not applicable to advanced melanoma, all other nevi (blue nevus, Reed's nevus, etc.) and non-melanocytic lesions (seborrhoeic warts, basal cell epithelioma, etc.). However, these lesions usually do not require DDA or follow-up for correct diagnosis.

Conflict of interest statement None.

References

1. Abbasi NR, Yancovitz M, Gutkowicz-Krusin D, Panageas KS, Mihm MC, Googe P et al (2008) Utility of lesion diameter in the clinical diagnosis of cutaneous melanoma. Arch Dermatol 144:469–474
2. Altamura D, Avramidis M, Menzies SW (2008) Assessment of the optimal interval for and sensitivity of short-term sequential digital dermoscopy monitoring for the diagnosis of melanoma. Arch Dermatol 144:502–506
3. Argenziano G, Fabbrocini G, Carli P, De Giorgi V, Delfino M (1997) Epiluminescence microscopy: criteria of cutaneous melanoma progression. J Am Acad Dermatol 37:68–74
4. Argenziano G, Kittler H, Ferrara G, Rubegni P, Malvehy J, Puig S et al (2010) Slow-growing melanoma: a dermoscopy follow-up study. Br J Dermatol 162(2):267–273
5. Argenziano G, Zalaudek I, Ferrara G, Johr R, Langford D, Puig S et al (2007) Dermoscopy features of melanoma incognito: indications for biopsy. J Am Acad Dermatol 56:508–513
6. Armitage P, Berry G (1987) Statistical methods in medical research. Blackwell Scientific Publications, Oxford
7. Barbini E, Cevenini G, Biagioli B, Scolletta S, Giomarelli P, Barbini P (2007) Predictive models of morbidity in intensive care unit—part I: model planning. Med Inform Decis Mak 7:1–16
8. Biagioli B, Scolletta S, Cevenini G, Barbini E, Barbini P, Giomarelli P (2006) A multivariate Bayesian model for assessing morbidity after coronary artery surgery. Crit Care 10:94
9. Bono A, Bartoli C, Cascinelli N, Lualdi M, Maurichi A, Moglia D et al (2002) Melanoma detection. A prospective study comparing diagnosis with the naked eye, dermatoscopy and telespectrophotometry. Dermatology 205:362–366
10. Bories N, Dalle S, Debarbieux S, Balme B, Ronger-Savlé S, Thomas L (2008) Dermoscopy of fully regressive cutaneous melanoma. Br J Dermatol 158:1224–1229
11. Bowling J, Argenziano G, Azenha A, Bandic J, Bergman R, Blum A et al (2007) Dermoscopy key points: recommendations from the international dermoscopy society. Dermatology 214:3–5
12. Burroni M, Corona R, Dell'Eva G, Sera F, Bono R, Puddu P et al (2004) Melanoma computer-aided diagnosis: reliability and feasibility study. Clin Cancer Res 10:1881–1886
13. Carli P, De Giorgi V, Giannotti B (2003) Why digital follow-up of dermoscopically equivocal pigmented lesions should be discouraged. Br J Dermatol 148:1272–1273
14. Carli P, Massi D, De Giorgi V, Giannotti B (2002) Clinically and dermatoscopically featureless melanoma: when prevention fails. J Am Acad Dermatol 46:957–959
15. Downing A, Yu XQ, Newton-Bishop J, Forman D (2008) Trends in prognostic factors and survival from cutaneous melanoma in Yorkshire, UK and New South Wales, Australia between 1993 and 2003. Int J Cancer 123:861–866
16. Dreiseitl S, Binder M, Hable K, Kittler H (2009) Computer versus human diagnosis of melanoma: evaluation of the feasibility of an automated diagnostic system in a prospective clinical trial. Melanoma Res 19:180–184
17. Ellenius J, Groth T (2000) Methods for selection of adequate neural network structures with application to early assessment of chest pain patients by biochemical monitoring. Int J Med Inform 57:181–202
18. Feit NE, Dusza SW, Marghoob AA (2004) Melanomas detected with the aid of total cutaneous photography. Br J Dermatol 150:706–714
19. Fukunaga K (1990) Introduction to statistical pattern recognition. Academic Press, Boston
20. Fuller SR, Bowen GM, Tanner B, Florell SR, Grossman D (2007) Digital dermoscopic monitoring of atypical nevi in patients at risk for melanoma. Dermatol Surg 33:1198–1206
21. Haenssle HA, Kaune KM, Buhl T, Thoms KM, Padeken M, Emmert S, Schön MP (2009) Melanoma arising in segmental nevus spilus: detection by sequential digital dermatoscopy. J Am Acad Dermatol 61:337–341
22. Haenssle HA, Krueger U, Vente C, Thoms KM, Bertsch HP, Zutt M et al (2006) Results from an observational trial: digital epiluminescence microscopy follow-up of atypical nevi increases the sensitivity and the chance of success of conventional dermoscopy in detecting melanoma. J Invest Dermatol 126:980–985
23. Haenssle HA, Vente C, Bertsch HP, Rupprecht R, Abuzahra F, Junghans V et al (2004) Results of a surveillance programme for patients at high risk of malignant melanoma using digital and conventional dermoscopy. Eur J Cancer Prev 13:133–138
24. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristics (ROC) curve. Radiology 143:29–36
25. Jain AK, Chandrasekaran B (1982) Dimensionality and sample size considerations in pattern recognition practice. In: Krishnaiah PR, Kanal LN (eds) Handbook of statistics, vol II. North Holland, Amsterdam, pp 835–855
26. Jennrich RI (1977) Stepwise discriminant analysis. In: Enslein K, Ralston A, Wilf HS (eds) Statistical methods for digital computers. Wiley, New York, pp 76–95
27. Kalkhoran S, Milne O, Zalaudek I, Puig S, Malvehy J, Kelly JW, Marghoob AA (2010) Historical, clinical, and dermoscopic characteristics of thin nodular melanoma. Arch Dermatol 146:311–318

28. Kittler H, Binder M (2001) Risks and benefits of sequential imaging of melanocytic skin lesions in patients with multiple atypical nevi. Arch Dermatol 137:1590–1595
29. Kittler H, Guitera P, Riedl E, Avramidis M, Teban L, Fiebiger M et al (2006) Identification of clinically featureless incipient melanoma using sequential dermoscopy imaging. Arch Dermatol 142:1113–1119
30. Kittler H, Pehamberger H, Wolff K, Binder M (2000) Follow-up of melanocytic skin lesions with digital epiluminescence microscopy: patterns of modifications observed in early melanoma, atypical nevi, and common nevi. J Am Acad Dermatol 43:467–476
31. Krzanowski WJ (1988) Principles of multivariate analysis: a user's perspective. Clarendon Press, Oxford
32. Lasko TA, Bhagwat JG, Zou KH, Ohno-Machado L (2005) The use of receiver operating characteristic curves in biomedical informatics. J Biomed Inform 38:404–415
33. Lindelöf B, Hedblad MA, Sigurgeirsson B (1998) Melanocytic naevus or malignant melanoma? A large-scale epidemiological study of diagnostic accuracy. Acta Derm Venereol 78:284–288
34. Makino E, Uchida T, Matsushita Y, Inaoki M, Fujimoto W (2007) Melanocytic nevi clinically simulating melanoma. J Dermatol 34:52–55
35. MATLAB (1996) The language of technical computing, using MATLAB, version 5. The MathWorks Inc., Natick
36. Mayer J (1997) Systematic review of the diagnostic accuracy of dermatoscopy in detecting malignant melanoma. Med J Aust 167:206–210
37. NIH Consensus Conference (1992) Diagnosis and treatment of early melanoma. JAMA 268:1314–1319, 30
38. Obuchowski NA, Lieber ML, Wians FH Jr (2004) ROC curves in clinical chemistry: uses, misuses, and possible solutions. Clin Chem 50:1118–1125
39. Obuchowski NA, Lieber ML (2002) Confidence bounds when the estimated ROC area is 1.01. Acad Radiol 9:526–530
40. Pizzichetta MA, Stanganelli I, Bono R, Soyer HP, Magi S, Canzonieri V et al (2007) Dermoscopic features of difficult melanoma. Dermatol Surg 33:91–99
41. Puig S, Argenziano G, Zalaudek I, Ferrara G, Palou J, Massi D et al (2007) Melanomas that failed dermoscopic detection: a combined clinicodermoscopic approach for not missing melanoma. Dermatol Surg 33:1262–1273
42. Risser J, Pressley Z, Veledar E, Washington C, Chen SC (2007) The impact of total body photography on biopsy rate in patients from a pigmented lesion clinic. J Am Acad Dermatol 57:428–434
43. Roseeuw D (2001) The invisible melanoma. J Eur Acad Dermatol Venereol 15:506–507
44. Roudier-Pujol C, Saiag P (1998) Malignant melanoma. Epidemiology, screening, diagnosis, evolution, prognostic clinical and histopathologic criteria. Rev Prat 48:1729–1735
45. Rubegni P, Burroni M, Andreassi A, Fimiani M (2005) The role of dermoscopy and digital dermoscopy analysis in the diagnosis of pigmented skin lesions. Arch Dermatol 141:1444–1446
46. Rubegni P, Burroni M, Cevenini G, Perotti R, Dell'Eva G, Barbini P et al (2002) Digital dermoscopy analysis and artificial neural network for the differentiation of clinically atypical pigmented skin lesions: a retrospective study. J Invest Dermatol 119:471–474
47. Rubegni P, Cevenini G, Burroni M, Perotti R, Dell'Eva G, Sbano P et al (2002) Automated diagnosis of pigmented skin lesions. Int J Cancer 101:576–580
48. Schiffner R, Schiffner-Rohe J, Landthaler M, Stolz W (2003) Long-term dermoscopic follow-up of melanocytic naevi: clinical outcome and patient compliance. Br J Dermatol 149:79–86
49. Schulman P (1984) Bayes' theorem—a review. Cardiol Clin 2:319–327
50. Skvara H, Teban L, Fiebiger M, Binder M, Kittler H (2005) Limitations of dermoscopy in the recognition of melanoma. Arch Dermatol 141:155–160
51. Soyer HP, Kerl H (1993) Surface microscopy of pigmented cutaneous tumors. Ann Dermatol Venereol 120:15–20
52. Strauss RM, Elliott F, Affleck P, Boon AP, Newton-Bishop JA (2007) A retrospective study addressed to understanding what predicts severe histological dysplasia/early melanoma in excised atypical melanocytic lesions. Br J Dermatol 157:758–764
53. Swerdlow AJ, Green A (1987) Melanocytic naevi and melanoma: an epidemiological perspective. Br J Dermatol 117:137–146
54. Urso C, Rongioletti F, Innocenzi D, Saieva C, Batolo D, Chimenti S et al (2005) Interobserver reproducibility of histological features in cutaneous malignant melanoma. J Clin Pathol 58:1194–1198