British Journal of Dermatology 2002; **147:** 481–486.

Clinical and Laboratory Investigations

Dermoscopic diagnosis by a trained clinician vs. a clinician with minimal dermoscopy training vs. computer-aided diagnosis of 341 pigmented skin lesions: a comparative study

D.PICCOLO, A.FERRARI, K. PERIS, R.DAIDONE, B.RUGGERI* AND S.CHIMENTI†

*Departments of Dermatology and *Internal Medicine and Public Health, University of L'Aquila, Via Vetoio – Coppito 2, 67100 L'Aquila, Italy*

†Department of Dermatology, University of Rome 'Tor Vergata', Italy

Accepted for publication 2 May 2002

Summary

Background In the last few years digital dermoscopy has been introduced as an additional tool to improve the clinical diagnosis of pigmented skin lesions.

Objectives To evaluate the validity of digital dermoscopy by comparing the diagnoses of a dermatologist experienced in dermoscopy (5 years of experience) with those of a clinician with minimal training in this field, and then comparing these results with those obtained using computer-aided diagnoses.

Methods Three hundred and forty-one pigmented melanocytic and non-melanocytic skin lesions were included. All lesions were surgically excised and histopathologically examined. Digital dermoscopic images of all lesions were framed and analysed using software based on a trained artificial neural network. Cohen's κ statistic was calculated to assess the validity with regard to the correct diagnoses of melanoma and non-melanoma.

Results Sensitivity was high for the experienced dermatologist and the computer (92%) and lower for the inexperienced clinician (69%). Specificity of the diagnosis by the experienced dermatologist was higher (99%) than that of the inexperienced clinician (94%) and the computer assessment (74%). Notably, computer analysis gave a higher number of false positives (26%) compared with the experienced dermatologist (0·6%) and the inexperienced clinician (5·5%).

Conclusions Our results indicate that analysis either by a trained dermatologist or an artificial neural network-trained computer can improve the diagnostic accuracy of melanoma compared with that of an inexperienced clinician and that the computer diagnosis might represent a useful tool for the screening of melanoma, particularly at centres not experienced in dermoscopy.

Key words: automatic diagnosis, computer-aided diagnosis, digital dermoscopy, neural network, pigmented skin lesions

The application and diffusion of dermoscopy, also known as dermatoscopy, epiluminescence microscopy and surface microscopy, has greatly improved the diagnostic accuracy of pigmented skin lesions (PSLs).[1–5] The recent development of sophisticated software programs that count and measure geometric and colorimetric parameters of the dermoscopic images has greatly improved the reproducibility of clinical assessment.[6] Furthermore, the combination of digital dermoscopy with a specific computer program, based on an artificial neural network (ANN), may represent an additional useful tool for the early diagnosis of melanoma, particularly for a clinician with minimal training in the field of PSLs. ANN is a special model of artificial intelligence based on the principles of neuronal propagation and processing used in several fields of medicine.[7,8] In dermoscopy, ANN can be applied for the pattern recognition of images.[9–13]

The first application of ANNs in the field of PSLs was reported in 1994 by Binder *et al.*[9] who performed a pilot study on 200 PSLs, achieving slightly lower

Correspondence: Ketty Peris.
E-mail: peris@univaq.it

values of sensitivity and specificity using a personal computer based on trained ANN compared with the values obtained by expert dermatologists. In 1998 the same authors conducted another study in order to classify PSLs dermoscopically using computerized image analysis and an ANN. Their results showed that the computerized system was able to identify 95% of PSLs automatically. The sensitivity and specificity values were 90% and 74%, respectively, when the discrimination was made between two categories of PSLs: common or dysplastic naevi in one category vs. melanoma in the other.[10]

In the same year, Seidenari *et al.*[11] reported the sensitivity and specificity values of an experienced and an untrained dermatologist compared with a computer analysis of 90 PSLs (59 benign PSLs and 31 melanomas). Sensitivity and specificity of the experienced observer were 81% and 95%, respectively, whereas lower values of sensitivity and specificity were obtained by an untrained observer (74% and 75%, respectively). Highest values of sensitivity were achieved by computer analysis (93%). The specificity value was not different from that obtained by the trained observer. In 1999, the same investigators assessed the efficacy of an automatic classifier, trained for 100% sensitivity using a subset of PSLs (59 naevi and 19 melanomas), on a test set including 365 naevi and 18 melanomas with Breslow thickness $< 0{\cdot}75$ mm. The specificity of the system reached 92%, whereas the sensitivity was 100%.[12]

In the present study we evaluated the diagnostic concordance on 341 PSLs between a dermatologist with 5 years of experience in the field of PSLs, specializing in dermoscopy, and a clinician with minimal training in dermoscopy. In addition, the results obtained from these two observers were compared with computer-aided diagnoses to evaluate the reliability and reproducibility of digital dermoscopy. Finally, we evaluated the differences between the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and reproducibility of the diagnosis made by investigators and the computer with regard to the age and phototype of the patients.

Materials and methods

Study population

Digital and dermoscopic images included in this study were from 341 PSLs [328 non-melanomas (comprising 316 naevi, seven dermatofibromas, three seborrhoeic keratoses and two angiomas) and 13 melanomas] that were observed and subsequently excised over a 6-month period from 289 patients (162 females and 127 males; mean age 33·6 years, range 3–83). All the lesions were excised because of equivocal dermoscopic findings or at the patient's request. All excised lesions were examined histopathologically by a dermatopathologist (S.C.).

Dermoscopic equipment

The equipment used for dermoscopic analysis consisted of a stereomicroscope with magnifications varying from $\times 6$ to $\times 40$ (Wild M-650; Leica Microscopy Systems Ltd, Heerbrugg, Switzerland), a high-resolution video camera (DXC 930P; Sony Corporation, Tokyo, Japan) 3 CCD, a personal computer with a Pentium 120 MHz processor with 16 MB RAM (Deskpro 4000; Compaq Computer Corporation, Houston, TX, U.S.A.), a high-resolution 20-inch colour monitor (PVM2053MD; Sony Corporation), and DEM-MIPS software (Digital Epi Microscopy Melanoma Image Processing Software; Biomips SRL, Siena, Italy). DEM-MIPS is based on an ANN trained with 100 PSLs (50 non-melanomas and 50 melanomas) and is designed to evaluate different colorimetric and geometric parameters of a lesion automatically in real time. All digital images of PSLs were collected in a Truevision Advanced Graphic Array format file with a size of 887 kB for each image. Digital dermoscopic images were framed at $\times 16$ magnification before analysis with DEM-MIPS. Cases were clinically and dermoscopically evaluated on a high-resolution colour monitor, in a random sequence, by both a trained dermatologist (5 years of experience) and a resident clinician with minimal training (6 months of experience, comprising 8 h of specialized training on three consecutive days and 2 h per week in the routine of dermoscopy) in the field of PSLs.

Statistical analysis

Sensitivity, specificity, PPV (i.e. the proportion of all cases diagnosed as melanomas that were histopathologically determined to be melanomas) and NPV (the proportion of all cases diagnosed as non-melanomas that were histopathologically determined to be non-melanomas) were calculated for both clinicians and for the computer. These values were adjusted for age and phototype of the patients. Reproducibility between trained clinician, resident clinician with minimal

training and the computer was evaluated with regard to age and phototype of the patients. Reproducibility was measured by Cohen's κ statistic. A κ-value of 1·0 indicates full agreement beyond chance. Values greater than 0·75 are generally considered excellent, values between 0·40 and 0·75 fair to good, and values less than 0·40 poor.[14] Statistical analysis was performed using Glim package, version 4·08 (Royal Statistical Society, London, U.K., 1992).

Results

Characteristics of patients with PSLs including melanomas are shown in Table 1. Based on histopathological analysis, the lesions included in this study consisted of 328 non-melanomas (96·2%) and 13 melanomas (3·8%). Non-melanoma skin lesions were classified as Clark naevi (281 of 328, 85·7%), Reed naevi (15 of 328, 4·6%), dermal naevi (12 of 328, 3·7%), blue naevi (seven of 328, 2·1%), dermatofibromas (seven of 328, 2·1%), seborrhoeic keratoses (three of 328, 0·9%), angiomas (two of 328, 0·6%) and one combined naevus (0·3%).

Results of sensitivity, specificity, PPV and NPV obtained by the trained dermatologist, inexperienced clinician and computer are shown in Table 2. The trained dermatologist identified two of 328 non-melanoma skin lesions (0·6%) as false positives and one of 13 melanomas (7·7%) as a false negative. The clinician with minimal training in dermoscopy identified 18 of 328 non-melanoma skin lesions (5·5%) as false positives and four of 13 melanomas (30·8%) as false negatives. Eighty-five of 328 non-melanoma skin lesions (25·9%) were detected as false positives by computer; the percentage of false negatives was exactly the same as that detected by the trained dermatologist.

Table 1. Characteristics of patients with pigmented skin lesions (PSLs)

	PSLs		Melanomas	
	n	%	*n*	%
Gender				
Male	153	44·9	9	69·2
Female	188	55·1	4	30·8
Age (years)				
< 25	75	22·0	–	–
25–39	162	47·5	2	15·4
> 39	104	30·5	11	84·6
Phototype				
I/II	107	31·4	3	23·1
III	144	42·2	4	30·7
IV/V	90	26·4	6	46·2

Reproducibility, measured using Cohen's κ statistic, and corresponding 95% confidence intervals, are shown in Table 3.

Discussion

Dermoscopy significantly improves the clinical diagnosis of PSLs, in particular the early detection of melanoma.[15–20] However, it has been previously demonstrated that the sensitivity of diagnostic accuracy is closely related to the specific experience of the investigator, and may be significantly lower in observers who are not formally trained.[21] In the last few years, digital image analysis has been introduced as a further refinement of 'classical' dermoscopy, using more sophisticated techniques. It allows quantification of colorimetric and geometric parameters of a PSL, introducing an objective element to the diagnosis and making diagnostic judgement more reproducible. Accuracy in diagnosis of PSLs has been further increased by using ANN, an approach involving artificial intelligence and information processing in computer science, allowing the input data of new lesions (clinical and dermoscopic criteria) to be processed on the basis of a training set database.[5–12] This integrated system should recognize the PSL, automatically extract features and use these criteria in training an ANN, which should then be capable of detecting and classifying a new PSL on the basis of the type of the input received in the training.[6,9–14,22,23]

Sensitivity and specificity values achieved by the computerized integrated system in discriminations of our PSLs (92% and 74%, respectively) were lower than those obtained by Seidenari *et al.*[11] (93% and 95%, respectively), but similar to the results achieved by Binder *et al.*[10] on two subsets of lesions including naevi and melanomas. Andreassi *et al.*[6] conducted a digital dermoscopic study on 147 PSLs (90 naevi and 57 melanomas) characterized by 'borderline' morphological features, using a combination of ANN with a multivariate stepwise discriminant analysis. The percentage of cases correctly classified was 85%, with a sensitivity of 88% and a specificity of 81%. These values are high considering the type of lesions used for the study. Recently, Rubegni *et al.*[13] obtained high values of sensitivity (96·5%) and specificity (95·4%) using ANN and stepwise discriminant analysis to differentiate between pigmented Spitz naevi and melanomas with similar clinical and dermoscopic features, thus demonstrating that ANN can be a useful tool for classifying PSLs.

Table 2. Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of diagnoses made by clinicians or computer, compared with histopathological diagnosis, based on age and phototype of the patients

		Age (years)[a]		Phototype			
Assessor	Parameter	25–39	> 39	I/II	III	IV/V	Total (95% CI)
Trained dermatologist	Sensitivity	1·00	0·91	0·67	1·00	1·00	0·92 (0·78–1·00)
	Specificity	0·99	0·99	0·99	0·99	1·00	0·99 (0·98–1·00)
	PPV	0·67	0·91	0·67	0·80	1·00	0·86 (0·67–1·00)
	NPV	1·00	0·99	0·99	1·00	1·00	0·99 (0·99–1·00)
Minimally trained clinician	Sensitivity	0·50	0·72	0·33	0·75	0·83	0·69 (0·44–0·94)
	Specificity	0·94	0·96	0·93	0·94	0·96	0·94 (0·92–0·97)
	PPV	0·09	0·67	0·12	0·27	0·62	0·33 (0·15–0·51)
	NPV	0·94	0·92	0·98	0·99	0·99	0·99 (0·92–0·98)
Computer	Sensitivity	0·50	1·00	0·67	1·00	1·00	0·92 (0·78–1·00)
	Specificity	0·72	0·75	0·73	0·76	0·71	0·74 (0·69–0·79)
	PPV	0·02	0·32	0·07	0·11	0·20	0·12 (0·06–0·19)
	NPV	0·99	1·00	0·99	1·00	1·00	0·99 (0·99–1·00)

CI, confidence interval. [a]Results regarding patients younger than 25 years are not reported because of lack of melanomas.

Table 3. Reproducibility of diagnoses made by clinicians or computer, based on age and phototype of the patients

	Kappa value[a]					
	Age (years)[b]		Phototype			
Comparison	25–39	> 39	I/II	III	IV/V	Total (95% CI)
TD vs. MTC	0·26	0·66	0·14	0·47	0·69	0·46 (0·26–0·65)
TD vs. computer	0·05	0·34	0·01	0·19	0·25	0·16 (0·01–0·32)
MTC vs. computer	0·08	0·32	0·10	0·19	0·26	0·19 (0·04–0·34)

CI, confidence interval; TD, trained dermatologist; MTC, minimally trained clinician. [a]$\kappa > 0{\cdot}75$, excellent agreement; $0{\cdot}40 < \kappa < 0{\cdot}75$, fair to good agreement; $\kappa < 0{\cdot}40$, poor agreement. [b]Results regarding patients younger than 25 years are not reported because of lack of melanomas.

The present study used an ANN trained to differentiate melanoma and non-melanoma skin lesions. We evaluated the diagnostic concordance between two clinicians with different degrees of experience in dermoscopy and a digital computer trained with an ANN. Results showed sensitivity and specificity of 92% and 99%, respectively, for the trained dermatologist, 69% and 94%, respectively, for the clinician with minimal training, and 92% and 74%, respectively, for the computer analysis.

Our results of a high sensitivity detected by the ANN are similar to those obtained by Bauer *et al.*[24] In contrast, we found a lower specificity (74%) compared with 97·8% specificity reported by Bauer *et al.* This discrepancy could be due to the different training used for ANN, such that our system was able to recognize a higher number of possible at-risk lesions.

The higher specificity value (99%) achieved by a dermatologist experienced in dermoscopy revealed the importance of dermoscopic training in avoiding a great number of unnecessary surgical excisions. Interestingly, the only melanoma misidentified as a naevus by the trained dermatologist was correctly determined by the computer.

Notably, computer analysis gave a higher number of false positive melanomas (26%) compared with the experienced dermatologist (0·6%) and the inexperienced clinician (5·5%). In addition, more false negatives (30·8%) were identified in the case of the clinician with minimal training. Thus, a well-trained ANN is capable of establishing a correct diagnosis of melanoma in a higher percentage of cases compared with a clinician with minimal training.

PPVs and NPVs were calculated assuming that the prevalence of melanoma (i.e. the probability of melanoma occurring in a subject, independent of the expert's diagnosis) was that obtained in our study (3·8% incidence of melanoma). Therefore, these values could change depending on the prevalence of melanoma in the study population.

In addition, we evaluated the differences between the sensitivity, specificity, PPV, NPV and reproducibility of the diagnosis made by human investigators and the computer with regard to the age and phototype of the patients. Interestingly, for the trained dermatologist there were no great differences in sensitivity and specificity with regard to the age and phototype of the patients. In contrast, sensitivity values of the inexperienced clinician and, surprisingly, of the computer, were remarkably higher when the patients were older than 39 years. It is conceivable that the inexperienced clinician was influenced by the age of the patient, whereas the trained dermatologist based the diagnosis exclusively on clinical and dermoscopic images. High values of sensitivity were calculated for both the trained and inexperienced observers as well as the computer for phototypes III–V. This might be explained by the fact that both clinicians in this study were accustomed to treating patients of Mediterranean descent, whose phototype is typically III or IV. However, this hypothesis should be confirmed in further studies.

Altogether, our results demonstrated that there was only just a fair to good agreement ($\kappa = 0{\cdot}46$) between the trained dermatologist and the clinician with minimal training in dermoscopy. In addition, there was a poor agreement ($\kappa = 0{\cdot}16$) between the trained dermatologist and computer as well as between the computer and the clinician with minimal training ($\kappa = 0{\cdot}19$). This low reproducibility is probably due to the great number of false positives of the computer analysis. Reproducibility between the trained dermatologist and the clinician with minimal training was better when the patient's age was >39 years and the phototype was III or IV.

In conclusion, our results indicate that analysis either by a trained dermatologist or ANN-trained computer can improve the diagnostic accuracy of melanoma although the specificity of computer diagnosis should be increased by a further training of ANN with the aim of reducing the number of false positives. This approach might represent a useful tool for the screening of melanoma, particularly at centres not experienced in dermoscopy. However, it is of paramount importance to clarify that computer analysis has been developed in order to assist and not to replace physicians in the diagnosis of PSLs and, indeed, the best diagnostic results were obtained when computer-derived data were managed by an experienced dermatologist.

Acknowledgments

This work was supported by the Ministry of the University for Scientific and Technological Research (MURST). The authors thank Barbara J.Rutledge for editing assistance.

References

1 Soyer HP, Smolle J, Kerl H, Stettner H. Early diagnosis of malignant melanoma by surface microscopy. *Lancet* 1987; **8**: 803.
2 Pehamberger H, Steiner A, Wolff K. *In vivo* epiluminescence microscopy of pigmented skin lesions. *J Am Acad Dermatol* 1987; **17**: 571–83.
3 Bahmer FA, Fritsch P, Kreusch J *et al.* Terminology in surface microscopy. *J Am Acad Dermatol* 1990; **23**: 1159–62.
4 Kenet RO, Kang S, Barney JK *et al.* Clinical diagnosis of pigmented lesions using digital epiluminescence microscopy: grading protocol and atlas. *Arch Dermatol* 1993; **129**: 157–74.
5 Argenziano G, Soyer HP. Dermoscopy of pigmented skin lesions – a valuable tool for early diagnosis of melanoma. *Lancet Oncol* 2001; **2**: 443–9.
6 Andreassi L, Perotti R, Rubegni P *et al.* Digital dermoscopy analysis for the differentiation of atypical nevi and early melanoma: a new quantitative semiology. *Arch Dermatol* 1999; **135**: 1459–65.
7 Asada N, Doi K, MacMahon H *et al.* Potential usefulness of an artificial neural network for differential diagnosis of interstitial lung disease: pilot study. *Radiology* 1990; **177**: 857–60.
8 Maclin PS, Dempsey J, Brooks J, Rand J. Using neural network to diagnose cancer. *J Med Syst* 1991; **15**: 11–19.
9 Binder M, Steiner A, Schwarz M *et al.* Application of an artificial neural network in epiluminescence microscopy pattern analysis of pigmented skin lesions: a pilot study. *Br J Dermatol* 1994; **130**: 460–5.
10 Binder M, Kittler H, Seeber A *et al.* Epiluminescence microscopy-based classification of pigmented skin lesions using computerized image analysis and an artificial neural network. *Melanoma Res* 1998; **8**: 261–6.
11 Seidenari S, Pellacani G, Pepe P. Digital videomicroscopy improves diagnostic accuracy for melanoma. *J Am Acad Dermatol* 1998; **39**: 175–81.
12 Seidenari S, Pellacani G, Giannetti A. Digital videomicroscopy and image analysis with automatic classification for detection of thin melanoma. *Melanoma Res* 1999; **9**: 163–71.
13 Rubegni P, Ferrari A, Cevenini G *et al.* Differentiation between pigmented Spitz nevus and melanoma by digital dermoscopy and stepwise logistic discriminant analysis. *Melanoma Res* 2001; **11**: 37–44.
14 Armitage P, Berry G. *Statistical Methods in Medical Research*, 2nd edn. Oxford: Blackwell Scientific Publishing, 1987.
15 Steiner A, Pehamberger H, Binder M, Wolff K. Pigmented Spitz nevi: improvement of the diagnostic accuracy by epiluminescence microscopy. *J Am Acad Dermatol* 1992; **27**: 697–701.
16 Pehamberger H, Binder M, Steiner A, Wolff K. *In vivo* epiluminescence microscopy: improvement of early diagnosis of melanoma. *J Invest Dermatol* 1993; **100**: 356–62.
17 Soyer HP, Smolle J, Leitinger G *et al.* Diagnostic reliability of dermoscopic criteria for detecting malignant melanoma. *Dermatology* 1994; **190**: 25–30.
18 Nachbar F, Stolz W, Merkle T *et al.* ABCD rule of dermoscopy. *J Am Acad Dermatol* 1994; **30**: 551–9.

19 Menzies S, Ingvar C, McCarthy WH. A sensitivity and specificity analysis of the surface microscopy features of invasive melanoma. *Melanoma Res* 1996; **6**: 55–62.
20 Argenziano G, Fabbrocini G, Carli P *et al.* Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. *Arch Dermatol* 1998; **134**: 1563–70.
21 Binder M, Schwarz M, Winkler A *et al.* Epiluminescence microscopy. A useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists. *Arch Dermatol* 1995; **131**: 286–91.
22 Sober AJ, Burstein JM. Computerized digital image analysis: an aid for melanoma diagnosis. *J Dermatol* 1994; **21**: 885–90.
23 Day GR, Barbour RH. Automated melanoma diagnosis: where are we at? *Skin Res Technol* 2000; **6**: 1–5.
24 Bauer P, Cristofolini P, Boi S *et al.* Digital epiluminescence microscopy: usefulness in the differential diagnosis of cutaneous pigmentary lesions. A statistical comparison between visual and computer inspection. *Melanoma Res* 2000; **10**: 345–9.