Melanoma Research, 1, pp. 231–236

Computer image analysis of pigmented skin lesions

A. Green,* N. Martin, G. McKenzie, J. Pfitzner, F. Quintarelli, B. W. Thomas, M. O'Rourke and N. Knight

Epidemiology Unit, Queensland Institute of Medical Research, Herston, Brisbane, Queensland 4029, Australia, Tel: (+61) 7 362 0234; Fax: (+61) 7 362 0111 (A. Green, N. Martin, G. McKenzie, J. Pfitzner, N. Knight; Centre for Medical and Health Physics, Queensland University of Technology, Brisbane, Australia (G. McKenzie, J. Pfitzner, F. Quintarelli, B.W. Thomas); Department of Surgery, Mater Misericordiae Hospital, Brisbane, Australia (M. O'Rourke).

To assist in the distinction of melanoma from benign pigmented lesions, an imaging system was developed, comprising a frame grabber, a microcomputer, a colour video camera and flash lighting with red, green and infrared filters. Over an 18-month period, video images of 70 unselected pigmented lesions for which complete diagnostic data were available, were successfully captured using the camera. Analysis software extracted features relevant to the size, colour, shape and boundary of each lesion, and these features were correlated with clinical and histological characteristics on which standard diagnoses of skin tumours are based. For discriminant analysis based on image analysis measurements, equal probabilities were assigned to three specified diagnostic groups, namely melanoma, naevi and 'other', and four of five melanomas were correctly classified when infrared data were included. However when infrared measurements were omitted, all five melanomas were correctly classified, and the overall accuracy of classification of pigmented lesions was 71%. This system holds promise as an aid in the clinical distinction of melanoma from benign pigmented skin lesions.

Key words: Early diagnosis, image analysis, melanoma, skin cancer.

Supported by the QIMR Trust Fund, the John P. Kelly Mater Research Foundation, the Mater Surgeons' Association, the Queensland Cancer Fund and the National Health and Medical Research Council of Australia. We are grateful to Darryl Collins for his early contribution, Dr Gerd Dowideit, the staff of the Mater Hospital Pathology Department and the clinicians and nursing staff for their generous assistance with this project.

*To whom correspondence should be addressed.

Introduction

Melanoma is one of the leading cancers in Australia,[1] and its incidence among many white-skinned populations around the world such as the United States[2] and northern European populations,[3] is rising rapidly. Its importance as a public health problem may increase further with the accelerating depletion of the ozone layer.[4]

Visual differentiation between various kinds of pigmented lesions of the skin is thought to be one of the more difficult diagnostic problems in medicine,[5] even among experienced clinicians, yet this ability is critical since patient survival depends on the early diagnosis and prompt excision of melanoma. As a corollary, removal of excessive numbers of benign naevi is a clinical burden and a source of concern and cost for patients. There are few data available on the accuracy of clinical diagnosis of melanoma. In one recently reported study of 10,735 skin tumours carried out over three periods between 1955 and 1982 in the Skin and Cancer Unit of New York,[6] it was found that among 277 melanomas, the positive predictive value ('diagnostic accuracy') in the best study period was 64%, with a sensitivity of 85%. Marshall[5] compared clinical and histological diagnoses of pigmented lesions in Cardiff, Wales, and found that among 63 lesions (44 suspected melanomas), local clinicians had a sensitivity of clinical diagnosis of 83%, specificity of 41%, and a diagnostic accuracy of 55%. In a selected series of 329 suspected melanomas submitted for frozen section in Brisbane between 1961 and 1973, 49% were histologically confirmed.[7]

Besides the immediate clinical requirement of accuracy in the diagnosis of pigmented lesions, research into the aetiology of melanoma is increasingly directed

towards the valid and reproducible assessment of these lesions given that the risk of melanoma is strongly related to the number of melanocytic naevi present on an individual.[8–10] Comparison of epidemiological studies of naevi in one population over time, or comparison of these risk factors between several different populations is hampered as there is no agreed method for discriminating between and counting the various types of pigmented lesions.[11]

Several methods have been proposed to answer the wide demand for more rigorous diagnosis of pigmented lesions. Marshall[5,12] investigated the use of photography to aid in the diagnosis of melanoma, but the techniques, although objective, required technical ability of a high order. Surface microscopy, as reported by Soyer *et al.*[13] allows an observer to examine pigmented lesions covered by a drop of immersion oil and a glass slide through a stereomicroscope, but this is time-consuming and finally relies on subjective interpretation. Dhawan[14,15] is developing an optical instrument called a 'Nevoscope' to enable thickness and three-dimensional size of a transilluminated pigmented lesion to be measured: this will assist in calculating the probability of malignancy.

We have developed software for an image processing system with the eventual aim of reliably and objectively characterizing malignant and benign pigmented lesions on the skin, including melanomas, naevi, freckles, pigmented basal cell carcinomas (BCC) and seborrhoeic keratoses in terms of their colour, colour variegation, shape, size, regularity of outline, and distinctness of boundary with surrounding skin. Using discriminant analysis, we have assessed the potential accuracy of using these image analysis measurements to classify a clinical series of pigmented lesions.

Methods

Subjects

Between February, 1989 and August, 1990, patients with pigmented lesions for excision were enrolled in the study through the surgery, dermatology and casualty departments of the Mater Misericordiae Hospital, Brisbane, with the approval of the institutional ethics committee. The treating doctor (in the majority of cases a surgeon or a dermatologist) completed a form giving clinical diagnosis of any pigmented lesions for excision, as well as standard clinical details, namely diameter (variable label, SIZE, mm); colour of lesion (COLOUR, 1. Uniform light brown 2. Uniform dark brown 3. Uniform black 4. Variegated); regularity of outline (OUTLINE, 1. Regular 2. Moderately irregular 3. Very irregular); diffuseness of edge (EDGE, 1. Clearly defined 2. Diffuse); and whether the lesion was palpable (PALPABLE, 1. Yes, 2. No). Similarly the hospital pathologists classified each study lesion according to predefined histological criteria for diagnosis, including size, histological type and, where appropriate, thickness, level, ulceration, mitotic rate, presence of regression, vascular or lymphatic invasion, margin of excision and whether tumour cells were present in the line of excision. Features extracted under image analysis for each lesion could thus be correlated with the clinical and histologic features on which diagnoses were based.

Imaging system

The prototype comprised a frame grabber; an IBM-compatible AT computer; a CCD colour video camera (modified National WV CD130) and custom-built flash lighting mounted on a tripod. The frame grabber was mounted in the computer and captured video images from the camera in 256 grey levels at 512 × 512 pixel resolution. As the frame grabber was monochrome, a means of obtaining sequential filtered images was required in order to obtain colour information. A bank of six electronic flash units was used for illumination because the units were cool and could be switched at video rates. Pairs of flashes (one for each video image frame) were covered with Kodak Wratten filters No 25 (red), No 58 (green) and No 87 (infrared) and fired sequentially, capturing three video image frames over 120 msec. Dedicated electronic circuitry under computer control synchronized the flash illumination with the image capture. A switching mechanism, triggered by the appropriate flash, enabled colour balancing of the camera. The digitized images were archived to floppy disk.

A standardised protocol for accumulating images at a known and reproducible focal distance was used. Each flash unit was oriented such that its centre of illumination coincided with the lesion 'in focus'. The image collection was timed, under computer control, to occur at the peak of the flash intensity. Any anomalies resulting from reflection when the flash illuminated the lesion, was eliminated by the processing software through comparison of the different fields contributing to the final image. Performance of these aspects of the system was verified using a test object.

The analysis software extracted the required fea-

tures from the filtered images. Patient movement in the 120 msec recording period resulted in imperfect registration of the three filtered images and therefore only one of the three images could be used to obtain the lesion boundary. The green-filtered image was chosen to threshold a boundary since it generally presented the greatest contrast between lesion and normal skin, and also provided the most accurate representation of the boundary shape. Perimeter tracing of the lesion using the red-filtered image gave variations which were inconsistent with the 'true' boundary, and so this was rejected as a suitable analytical tool for boundary determination. The infrared image was not used for determining the lesion boundary because the lesion was not observed in a number of images at this wavelength. A statistically based algorithm was used tc determine the thresholding level and an eight neighbourhood contour follower was used to trace the boundary of the lesion.

Image analysis

The features extracted for image processing were the means and variances of the grey level values from red (REDMN, REDV), green (GREENMN, GREENV), and infrared (IRMN, IRV) filtered images within the boundary defined by the green-filtered image; the lesion edge generated by determining the mean and variance (EDGEMN, EDGEV) of the gradient between regions comprising a lesion component and an adjacent skin component; area of lesion (AREA); perimeter (PERIM); and fragmentation index (FRAG, 4πAREA/PERIM^2), which has a value of unity if a lesion is a perfect circle and smaller value if it is irregular.

Data analysis

Clinical, histological and image analysis measurements were subjected to discriminant analysis using SPSS-PC,[16] in order to see whether three classes of lesion, namely melanoma, melanocytic naevi and other pigmented lesions, could be distinguished on the basis of linear combinations of the measurements. Discriminant analysis was performed firstly with all available computer image analysis (CIA) measurements and then using a stepwise process of inclusion and elimination to decide which variables were most critical by minimizing Wilks' lambda. Analyses were performed with all 11 CIA measurements, and then omitting the two infrared measurements, since we were particularly interested in the value of IR imaging. We also added the five clinical measurements to see whether they added any greater precision to lesion classification beyond that provided by image analysis. Classification following discriminant analysis depends critically on the prior probabilities assigned to the groups one is attempting to distinguish. Although it could be argued that naevi and other benign lesions are much more common than melanomas, we chose to give the three groups equal probability on the basis that this weights the odds towards a false positive rather than a false negative diagnosis of melanoma, which would be preferred in the clinical situation.

Results

Study sample

There were 89 successfully imaged lesions which were subsequently excised from an unselected group of 81 patients (45 female) with a median age of 32 years. Complete clinical and histologic information was available for 62 of these lesions: another eight lesions had clinical diagnoses assigned (all clearly benign) in the absence of available histology reports. Approximately 80% of the 70 lesions studied were removed from the trunk, 10% from the face or neck, and 10% from the limbs. Five of the lesions were histologically diagnosed as melanoma, 53 were melanocytic naevi (two of which were dysplastic) and 12 were miscellaneous pigmented lesions comprising seven seborrhoeic keratoses, two BCC, two skin tags and one lentigo. Of the five melanomas, three were diagnosed by the treating doctor, one had been clinically diagnosed as a dysplastic naevus and one as a pigmented BCC. Among the 53 naevi, 32 had been correctly diagnosed as common naevi, 17 were clinically diagnosed as dysplastic naevi and two as melanomas.

Correlations of clinical and imaged features

It is interesting to examine the correlations of clinical ratings with their corresponding CIA measurements. Clinical size measurement was highly correlated with area and perimeter as expected, whereas clinically assessed colour was not significantly correlated with CIA colour variables (REDMN, GREENMN, IRMN) (Table 1). Similarly, there was no association between clinical grading of a lesion's regularity of outline and computed fragmentation index, or between clinical

Table 1. Correlations (×100) between clinical ratings (variables 1–5) and image analysis measurements (variables 6–16) of 70 pigmented lesions

1.	SIZE	•														
2.	COLOUR	31	•													
3.	OUTLINE	35	52	•												
4.	EDGE	7	34	29	•											
5.	PALPABLE	−4	6	18	−5	•										
6.	REDMN	6	−15	15	2	−19	•									
7.	GREENMAN	12	−12	18	6	−24	90	•								
8.	IRMN	9	−19	13	−2	−9	86	76	•							
9.	REDV	7	13	−17	0	7	−29	−28	−5	•						
10.	GREENV	8	−2	0	0	−8	35	51	39	39	•					
11.	IRV	17	22	−12	4	−17	−55	−40	−53	55	9	•				
12.	EDGEMN	8	9	2	−12	6	4	18	0	−1	23	12	•			
13.	EDGEV	11	0	−3	0	5	4	20	12	17	41	11	77	•		
14.	AREA	65	8	1	−1	−16	0	8	−3	20	21	38	12	17	•	
15.	PERIM	63	21	15	9	−13	14	13	9	19	16	23	15	15	77	•
16.	FRAG	−4	−24	−14	−21	6	−18	−7	−3	0	9	4	−15	−10	−2	−53
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Two-tail significance levels of 5%, l% and 0.1% are achieved by $r > 0.23, 0.30, 0.38$, respectively.

assessment of diffuseness of edge and the imaged features EDGEMN and EDGEV (Table 1).

Lesion classification using image analysis

The value of each of the 11 CIA variables for distinguishing the three diagnostic groups of lesions is shown in Table 2. The F ratio and its associated probability testing the equality of the three group means for each variable is listed, along with Wilks' lambda (or U statistic) which is the ratio of the within-groups sum of squares to the total sum of squares. If the group means are equal this will have a value of 1, while small values indicate that group means are different. Inspection of Table 2 revealed that it was the colour means which promised to be most useful in discriminating between groups in the order IR, red and green, in addition to the IR and green variances. The only shape variable of potential use appeared to be the fragmentation index. Physicians' rating of the diffusiveness of edge appeared to be the only clinical feature of even modest discriminatory value (Table 2).

The classification of lesions using discriminant functions based on all 11 measurements is shown in Table 3(a). Stepwise analysis however, revealed that virtually the same discrimination could be achieved with only four variables – IRMN, FRAG, REDV and IRV. Using the two functions based on these variables, each of the three groups was significantly different from the other two at the 1% level of significance or higher. To consider their role in the discriminant function the two infrared measures IRMN and IRV were omitted from the analyses. Classification results from the analysis based on all nine measurements are shown in Table 3(b), and although overall classification accuracy dropped from 76% to 71%, the classification of the melanomas actually improved, with all five being correctly classified compared with only four when the IR measurements were used. Stepwise analysis showed that the five most important variables for discrimination were (in order) REDMN, FRAG, REDV, area and perimeter. Of the two functions defined with these variables,

Table 2. Wilks' Lambda (U-statistic) and univariate F-ratio testing equality of diagnostic group means for each variable.

	Lambda	$F_{2,67}$	P
SIZE	0.98	0.58	0.561
COLOUR	0.97	0.87	0.423
OUTLINE	0.99	0.25	0.777
EDGE	0.96	1.16	0.318
PALPABLE	0.97	0.95	0.388
REDMN	0.84	6.00	0.004
GREENMN	0.88	4.41	0.015
IRMN	0.72	12.76	0.000
REDV	0.97	0.97	0.383
GREENV	0.96	1.11	0.335
IRV	0.91	3.08	0.052
EDGEMN	0.98	0.38	0.679
EDGEV	0.99	0.28	0.753
AREA	0.98	0.65	0.522
PERIM	0.98	0.53	0.587
FRAG	0.90	3.39	0.039

Table 3. Classification results of discriminant analysis based on (a) all image analysis measurements, (b) omitting infrared measurements

(a) All 11 image analysis measurements

Actual group	No. of cases	Predicted group membership		
		Melanoma	Naevus	Other
Melanoma	5	4	1	0
Naevus	53	4	43	6
Other lesion	12	2	4	6

76% of cases correctly classified

(b) Omitting infrared mean and variance

Actual group	No. of cases	Predicted group membership		
		Melanoma	Naevus	Other
Melanoma	5	5	0	0
Naevus	53	8	40	5
Other lesion	12	3	3	6

73% of cases correctly classified

the second was not quite significant ($p = 0.06$), and while the melanoma and naevus comparison and the naevus and 'other' comparison were each significant at the 2% level, the melanoma and 'other' groups were not significantly different. Classification with these reduced functions had about the same accuracy as involving all variables, but one of the melanomas was now misclassified.

Discriminant analysis with the five clinical variables alone resulted in classification accuracy of less than 50% – only slightly better than chance, and adding the clinical variables to the image analysis resulted in no significant improvement in overall discrimination between lesion groups, although 'other' lesions were more often correctly classified.

Discussion

We have described the prototype of an imaging system to assist in the diagnosis of pigmented lesions, and in particular, the distinction of melanoma from benign pigmented lesions. Based on analysis of histological parameters considered to be diagnostic, this system appears to provide useful measurements of the critical observable features of melanoma and could therefore aid, in non-invasively distinguishing malignant pigmented lesions from benign lesions. Because of the equal prior probabilities that were set in these analysis, the test tended to be more sensitive than specific in diagnosing melanoma, which is desirable. However these assumptions need further examination and refinement before application of the system in various clinical settings.

Previous work has suggested that the level of IR reflectance from a suspicious lesion may be associated with malignancy.[12] Although the infrared reflectance level appeared to the most powerful discriminatory CIA variable in distinguishing between diagnostic groups, this was not confirmed in the actual classification of the small number of melanomas in the study series: all five melanomas were correctly classified without inclusion of IR measurements (compared with four out of five when IR variables were included). Apart from IR, the most important discriminating features were the red level (mean and variance), irregularity of border, and size of the lesion, which cohere with the standard description[17] used to alert clinicians to the possibility of malignancy: colour variegation in a large, irregularly-shaped lesion.

Image analysis appeared to discriminate between melanomas and benign lesions better than the recorded clinical features, though in these analyses we accounted for neither the stated clinical diagnosis nor the degree of certainty the doctor felt in making that diagnosis. Dhawan in his system[15] advocates integration of such objective data with prognostic features from the patient's history, and using a knowledge base of case studies, calculating and issuing a score of probability of malignancy. Ultimately we envisage this imaging system as a useful diagnostic adjunct to practitioners assessing the likelihood of malignancy, rather than a complete diagnostic tool.[15]

Several problems prevented analysis of many of the images initially recorded in this study, including cases in which the software failed to separate the lesion from the surrounding skin due to either low contrast in the green filtered images (on which the algorithm segmented), or failure to recognise hairs in the lesion causing the tracing of an illogical boundary; cases where equipment size prevent the imaging of awkwardly-situated lesions e.g. behind the ear; cases of inadequately pigmented lesions being submitted for imaging e.g. skin tags and BCC; or cases in which the flash lighting failed or disc errors in the data files occurred. However we regard the preliminary findings as encouraging enough to develop the CIA system by improving the software, and modifying the hardware using commercially available items, aiming to produce a portable computer image analysis system to enhance the clinical detection of melanoma.

References

1. Giles G, Armstrong B, Smith L. *Cancer in Australia 1982.* National Cancer Statistics Clearing House. Sci Public No 1, 1987.
2. Glass AG, Hoover RN. The emerging epidemic of melanoma and squamous cell skin cancer. *J Am Med Assoc* 1989; **262:** 2097–2100.
3. Osterlind A, Hou-Jensen K, Moller Jensen O. Incidence of cutaneous malignant melanoma in Denmark 1978–1982. Anatomic site distribution, histologic types, and comparison with non-melanoma skin cancer. *Br J Cancer* 1988; **58:** 385–391.
4. Editorial. Ozone depletion quickens. *Lancet* 1991; **1:** 1132–1133.
5. Marshall RJ. Evaluation of a diagnostic test based on photographic photometry of infrared and ultraviolet radiation reflected by pigmented lesions of skin. *J Audiovisual Med* 1980; **3:** 94–98.
6. Grin CM, Kopf AW, Welkowich B, *et al.* Accuracy in the clinical diagnosis of malignant melanoma. Arch Dermatol 1990; **126:** 763–766.
7. Little JH, Davis NC. Frozen section diagnosis of suspected malignant melanoma of the skin. *Cancer* 1974; **34:** 1163–1172.
8. Green A, MacLennan R, Siskind V. Common acquired naevi and the risk of malignant melanoma. *Int J Cancer* 1985; **35:** 297–300.
9. Holly EA, Kelly JW, Shpall SN, *et al.* Number of melanocytic naevi as a major risk factor for malignant melanoma. J Am Acad Dermatol 1987; **17:** 459–468.
10. English JSC, Swerdlow AJ, MacKie RM, *et al.* Relation between phenotype and banal melanocytic naevi. *Br Med J* 1987; **294:** 152–154.
11. Green A, Swerdlow AJ. Epidemiology of melanocytic naevi. *Epidemiol Rev* 1989; **11:** 204–221.
12. Marshall R. Infrared and ultraviolet reflectance measurements as an aid to the diagnosis of pigmented lesions of skin. *J Audiovisual Med* 1981; **4:** 11–14.
13. Soyer HP, Smolle J, Hodl S, *et al.* Surface microscopy: a new approach to the diagnosis of cutaneous pigmented tumors. *Am J Dermatopath* 1989; **11:** 1–10.
14. Dhawan AP. Early detection of cutaneous malignant melanoma by three-dimensional nevoscopy. *Comp Meth Prog Biomed* 1985; **21:** 59–68.
15. Dhawan AP. An expert system for the early detection of melanoma using knowledge-based image analysis. *Anal Quant Cytol Histol* 1988; **10:** 405–416.
16. Norusis MJ. *SPSS/PC+ Advanced Statistics 4.0.* Chicago, SPSS Inc.
17. Davis NC. Malignant melanoma: the Australian contribution. *Aust NZ J Surg* 1988; **58:** 605–617.

(Received 25 July 1991; accepted 12 September 1991)