Computer image analysis in the diagnosis of melanoma

Adèle Green, MD,[a] Nicholas Martin, PhD,[a] John Pfitzner, MSc,[a]
Michael O'Rourke, FRACS,[b] and Ngaire Knight, RN[a]
Brisbane, Queensland, Australia

Background: It is often difficult to differentiate early melanoma from benign pigmented lesions of similar clinical appearance.
Objective: Our purpose was to develop a computer image analysis system that has the potential for use as an adjunct to the clinical distinction of melanoma from less serious pigmented lesions.
Methods: The system, consisting of a hand-held device incorporating a color video camera and color frame grabber mounted in a microcomputer, was used in a pigmented lesion clinic. Analysis software extracted features relevant to the size, color, shape, and boundary of each lesion, and these features were correlated with clinical and histologic characteristics on which standard diagnoses of skin tumors are based. For discriminant analysis based on image analysis measurements, equal prior probabilities were assigned to two specified diagnostic groups, namely melanoma and "other pigmented lesions," most of which were melanocytic nevi.
Results: In a 20-month period, video images of 164 unselected pigmented lesions for which complete diagnostic data were available were successfully captured using the camera. Sixteen of 18 melanomas, and 89% of pigmented lesions overall, were correctly classified by the image analysis system, compared with 83% based on clinical gradings of lesion characteristics.
Conclusion: Computer image analysis has the potential to provide a valuable diagnostic aid that could enable clinicians to make highly sensitive and specific diagnoses of early, curable melanoma.
(J Am Acad Dermatol 1994;31:958-64.)

From the Epidemiology Unit, Queensland Institute of Medical Research,[a] and the Department of Surgery, Mater Misericordiae Hospital.[b]

Supported by the J. P. Kelly Mater Research Foundation, the Queensland Cancer Fund, and the National Health and Medical Research Council of Australia.

Reprint requests: Adèle Green, MD, Epidemiology Unit, Queensland Institute of Medical Research, Herston, Brisbane, Queensland 4029, Australia.

0190-9622/94 $3.00 + 0 **16/1/57266**

There is increasing interest in the possibility that new imaging technologies may be used to enhance the diagnosis of cancer of various types. Cutaneous melanoma is unique among cancers in the sense that it is readily accessible and highly contrasted with the surrounding skin. Theoretically, computer imaging techniques could be applied in the diagnosis of melanoma directly and relatively easily. In comparison with other cancers such as breast cancer in which the

aim of imaging is to detect the existence of a tumor, the clinical challenge with melanoma, particularly early melanoma, is to differentiate it accurately from benign pigmented lesions of similar clinical appearance. Many early melanomas are excised with the diagnosis unsuspected,[1] and this raises the question of the proportion of invasive, and potentially invasive, melanomas that may be unsuspected but are not excised at this early stage.

In addition to computer image analysis (CIA), several other methods have been explored in the search to facilitate the diagnosis of pigmented lesions. In particular, the technique of epiluminescence microscopy has received much attention. It involves the use of oil immersion to make pigmented structures beneath the skin surface that cannot be seen by the naked eye, accessible to microscopic examination.[2] Preliminary data indicate that inspection of the pigment structure at the levels of the epidermis, dermoepidermal junction, and papillary dermis may reveal various patterns in the pigment network that are specific to malignant versus benign lesions.[3] The technique relies heavily on the expertise of the observer in the interpretation of the findings.

We have developed software for an image processing system with the eventual aim of reliably and objectively distinguishing melanomas from less serious pigmented lesions on the skin, such as nevi, freckles, pigmented basal cell carcinomas, and seborrheic keratoses. Potential diagnostic characteristics include color, variegation, shape, size, regularity of outline, and distinctness of boundary with surrounding skin. With discriminant analysis, we have assessed the accuracy of image analysis measurements of these visual features in classifying a series of more than 150 pigmented lesions.

METHODS

Subjects

Between August 1990 and April 1992, patients with pigmented lesions for excision were enrolled in the study, as in an earlier study.[4] The treating physician completed a form giving the clinical diagnosis, as well as standard clinical details, namely diameter (variable label, SIZE: millimeters); color of lesion (COLOR 1: uniform light brown, blue, or red; 2: uniform dark brown; 3: variegated; 4: black); 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 lesion according to standard histologic criteria including size, histologic type, and, where appropriate, thickness, Clark level, ulceration, mitotic rate, presence of regression, vascular or lymphatic invasion, margin of excision, and whether tumor 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 the diagnosis was based.[4]

Imaging system

To provide a less cumbersome system than the previously described prototype,[4] a hand-held device encasing a minature CCD color camera (Panasonic WV-CD1E) and annular light source (model KL1500 [Schott Glaswerke, Wiesbaden, Germany] with a halogen bulb and operating at a color temperature of 3000 K) was custom built. It has a field of view of approximately 30 × 20 mm at a fixed focal length with a magnification of ×6. An SVHS video cassette recorder (Panasonic AG 7330) with a horizontal resolution of approximately 240 lines and with dedicated electronic control circuitry, records in 8-second segments from this device. Images for analysis were captured from the replayed video tape, with a VISIONplus-AT color frame grabber mounted in an IBM-compatible AT computer. The digitized images were 24 bits per pixel (8 bits each for red, green, and blue (RGB) planes) with a 512 × 512 pixel resolution. Image capture and display were performed on the AT computer and image processing on an HP9000. The analysis software extracted the required features from the images. Color calibration was achieved by averaging the RGB values for 9 gray level standards (Kodak) taken from 10 consecutive images; all subsequent RGB values were mapped accordingly.

Segmentation of images

The algorithm used for segmentation of the images was an adaption of gray level thresholding to three-dimensional color space, that is, RGB level intensities. The algorithm relied on the images having two main regions that differed significantly in color but had similar color values within each region. Therefore two distinct regions in color space were obtained, with the transition from skin to lesion between them. To reduce the determination of color threshold to one dimension, the color values were mapped to a vector between the two main regions. A histogram was thus obtained on which gray level thresholding techniques were used, that is, the turning point or valley between two peaks was the threshold. Because this method used color differences, it was applicable to a variety of skin and lesion colors. To gauge the practical efficacy of the thresholding algorithm, the repeatability of the imaging technique was assessed on four replicate images obtained for a random subset of 57 lesions.

Image analysis

The features extracted for image processing were the means and standard deviations of values from red (RED-

Table I. Correlations between clinical and image analysis measurements in 164 lesions*

	r_i*	1	2	3	4	5	6	7	8	9	10
1. SIZE	•	•									
2. COLOR	•	12	•								
3. OUTLINE	•	35	17	•							
4. EDGE	•	13	11	29	•						
5. PALPABLE	•	−06	06	17	11	•					
6. REDMN	98	−04	−23	−01	19	10	•				
7. GREENMN	96	−03	−15	−00	18	12	95	•			
8. BLUEMN	95	−02	−09	−01	19	13	91	97	•		
9. REDSD	80	16	13	04	27	−04	−00	01	07	•	
10. GREENSD	75	20	08	05	22	−08	23	27	30	87	•
11. BLUESD	78	28	14	13	24	−14	18	22	28	75	92
12. RGRADMN	98	−08	17	−12	−26	−18	−60	−55	−53	−00	−04
13. GGRADMN	98	−03	15	−11	−27	−19	−55	−52	−50	00	−00
14. BGRADMN	98	−01	14	−10	−27	−19	−52	−49	−47	01	02
15. RGRADSD	94	34	16	16	−11	−19	−39	−35	−32	14	14
16. GGRADSD	93	39	15	19	−11	−17	−33	−29	−27	14	16
17. BGRADSD	93	40	15	20	−11	−16	−33	−28	−26	13	16
18. AREA	99	85	14	33	07	−10	−14	−12	−11	20	20
19. PERIM	95	78	12	42	08	−03	−05	−02	−03	07	05
20. FRAG	91	−39	−03	−42	−15	−10	−13	−13	−11	09	15
21. AREACH	37	14	08	26	34	18	53	55	56	06	14
22. SHAPECH	12	16	−05	03	10	−00	29	30	25	08	16

Two-tail significance tests of 5%, 1%, and 0.1% are achieved by $r > 0.15$, 0.20, and 0.26, respectively.
*The left hand column gives the estimated common inter-item correlation (r_i x.) based on four replicate images of 57 lesions.

MN, REDSD), green (GREENMN, GREENSD) and blue (BLUEMN, BLUESD) planes within the lesion boundary; the means and standard deviations of the gradient at each boundary pixel in the red (RGRADMN, RGRADSD), green (GGRADMN, GGRADSD), and blue (BGRADMN, BGRADSD) planes (the gradient at each boundary pixel was determined from the averages of skin and lesion pixels in a 24-neighborhood mask); area of lesion (AREA); and perimeter (PERIM). Other calculated values were the fragmentation index (FRAG, $4\pi AREA/PERIM^2$), which has a value of unity if a lesion is a perfect circle and smaller values if it is irregular; change in area (AREACH) when the color thresholds were reduced by 10%, as a broad indication of the gradient of color change around the lesion boundary; and shape change (SHAPECH), which was the ratio of fragmentation indices obtained at the two threshold levels (calculated threshold and 90% of calculated threshold).

Data analysis

Clinical, histologic, and image analysis measurements were subjected to discriminant analysis by means of SPSS-PC,[5] to determine whether histologically diagnosed melanomas could be distinguished from melanocytic nevi and the other pigmented lesions on the basis of linear combinations of the measurements. Analyses were performed with all 17 CIA measurements. We also added the five clinical measurements to determine whether they improved the precision of lesion classification by image analysis alone. Classification after discriminant analysis depends critically on the prior probabilities assigned to the groups one is attempting to distinguish. As we have done previously,[4] we chose to give the two groups (melanoma and other pigmented lesions) equal probability despite the relative ubiquity of the latter, 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. The measurements extracted from the 57 replicated images were subjected to repeated measures analysis with the SPSS-PC RELIABILITY procedure.

RESULTS

Study sample

A total of 199 images were collected. Three of these contained two pigmented lesions of interest and one contained three, making a total of 204 lesions available for analysis. The same software factors were used for the analysis of all lesions except for the images with multiple lesions in which bounds were placed around the individual lesions and seven other lesions in which the boundaries had to be manually drawn. This had to be done because of the

11	12	13	14	15	16	17	18	19	20	21	22
•											
02	•										
06	99	•									
10	98	99	•								
24	62	66	67	•							
25	56	61	62	99	•						
26	56	60	62	98	99	•					
27	05	10	12	56	60	61	•				
11	−14	−10	−08	41	46	47	86	•			
13	49	48	47	05	01	01	−29	−62	•		
16	−47	−45	−44	−12	−06	−05	10	26	−26	•	
11	−21	−19	−18	−15	−13	−13	11	07	−06	15	•

highly variable pigmentation of the lesion and/or the surrounding skin (caused by gross mottling from sun damage) or because the boundary of the image of a large lesion abutted the edge of the field of view. Usable values were obtained from 171 of the 204 lesions. Reasons for exclusion were that the lesion was too big (four cases), or obscured by hairs (eight) or surgeons' pen marks (two); in nine cases the software was unable to contend with the lesion characteristics, mainly because the lesion was too light or too fragmented; another 10 cases were missed because of avoidable operator error. Histopathologic examination was not available for another seven lesions. The final total of 164 lesions with complete clinical, histopathologic, and imaging data came from 129 patients (74 female) with a mean age of 36 years (range, 6 to 87 years).

Approximately 66% of the 164 lesions studied were removed from the trunk, 10% from the face or neck, and 24% from the limbs; 38 of the 164 lesions were clinically diagnosed as melanoma. On the basis of histologic diagnosis, there were 18 melanomas, 3 lentigo maligna, 128 melanocytic nevi (7 of which were dysplastic), and 15 miscellaneous pigmented lesions including seborrheic keratoses, basal cell carcinomas, and lentigines. Of the 18 melanomas, 14 of which were level I or II, 15 were diagnosed by the treating physician, whereas two had been clinically diagnosed as dysplastic nevi and one as a benign nevus. Of the 128 histologically diagnosed nevi, clinical diagnoses included benign nevi (78) and melanoma (12), as well as dysplastic nevi (35) of which 3 were histologically confirmed; the remainder were compound (28), junctional (2), and intradermal (2).

Repeatability measures

Among the set of replicate images obtained (4 replicate images of each of 57 random lesions), the estimated common inter-item correlations (r_i) or repeatabilities of the image analysis measures were calculated (shown in the left-hand column of Table I). With the exception of the two change measures (AREACH, SHAPECH), these were all high with the standard deviation measures less repeatable than their mean equivalents as might be expected.

Correlations of clinical and imaged features

As in the previous series[4] we examined the correlations of clinical ratings with their corresponding CIA measurements (Table I). Clinical size measurement (SIZE) was highly correlated with AREA (0.85) and perimeter (PERIM) (0.78) as expected. Clinically assessed COLOR was corre-

Table II. Pooled-within-groups correlations (ordered by size) between discriminating variables and the canonical discriminant function

Variable	r
AREA	0.71
PERIM	0.57
BGRADSD	0.45
BLUESD	0.45
GGRADSD	0.45
RGRADSD	0.41
GREENSD	0.33
REDSD	0.31
FRAG	−0.17
AREACH	0.09
SHAPECH	0.07
BGRADMN	0.14
REDMN	−0.10
GGRADMN	0.14
RGRADMN	0.11
GREENMN	0.01
BLUEMN	0.03

lated with CIA mean color variables (REDMN, GREENMN) and all the gradient variables (Table I). Clinical grading of a lesion's regularity of outline (OUTLINE) was significantly correlated with both perimeter (0.42) and computed fragmentation index (FRAG, −0.42), and clinical assessment of diffuseness of edge (EDGE) correlated predictably with the imaged color variance measures (0.22 to 0.27), the three gradient means (−0.26 to −0.27) and AREACH (0.34). There was no CIA analogue of clinical palpability.

Lesion classification by means of image analysis

The value of each of the 17 CIA variables for distinguishing the two diagnostic groups of lesions (melanoma and other) is shown in Table II, which lists the correlations of each variable with the canonical discriminant function. Use of this discriminant function resulted in correct classification of 16 of 18 melanomas and of 89% lesions overall; 16 other lesions were incorrectly classified as melanomas (Table III). Table II reveals that it was the size variables area and perimeter, followed by the three gradient SDs, the three color SDs, and FRAG and AREACH that were most useful in discriminating between groups. Not useful were the color and gradient mean variables, or SHAPECH.

For comparison, we also performed discriminant analysis with the five clinical ratings alone. Correlations with the discriminant function were 0.76 for SIZE and OUTLINE, COLOR (0.36), EDGE (0.32), and PALPABLE (−0.02). On the basis of this clinical measures function, 83% of lesions were properly classified (Table III), with two melanomas and 26 other lesions incorrectly classified. Combination of both clinical and image analysis measurements in the same discriminant function analysis resulted in no improvement over either set separately.

Table III. Classification results of discriminant analysis based on all 17 image analysis measurements and clinical ratings

Actual group	No. of cases	Predicted group membership: Melanoma	Predicted group membership: Other
*All 17 image analysis measurements**			
Melanoma	18	16	2
Other lesion	146	16	130
Clinical ratings†			
Melanoma	18	16	2
Other lesion	146	26	120

*89% of cases correctly classified.
†83% of cases correctly classified.

Because we particularly wanted to avoid false-negative results, we examined the two melanomas misclassified by the CIA discriminant function, the two cases misclassified by the clinical measures function and the three by clinical diagnosis. The same two melanomas, both in situ melanomas, were misclassified by all three diagnostic functions. One was a dark brown, impalpable lesion, 5 mm in diameter, removed from the leg of a 43-year-old woman. The clinical diagnosis was a benign nevus; the histologic diagnosis was a superficial spreading melanoma, Clark level I, with a measured thickness of 0.3 mm and with an adjacent junctional nevus. The other melanoma incorrectly classified by CIA and by the clinical measures function occurred on the side of the foot of a 24-year-old woman and was clinically assessed as a dysplastic nevus. The histologic diagnosis was acral lentiginous melanoma, level I, 0.2 mm thick. In addition, there was one lesion for which the clinical diagnosis differed from that of the pathologist: a 12 mm impalpable variegated lesion on the chest of an 18-year old man. The clinical diagnosis was dysplastic nevus, and the pathologic diagnosis was superficial spreading melanoma, level II, with a thickness of 0.3 mm.

DISCUSSION

We previously described the prototype of an imaging system that had the potential to assist in distinguishing melanoma from less serious pigmented lesions.[4] On the basis of analysis of histologic features considered to be diagnostic, the prototype imaging system appeared to provide useful measurements of critical observable features of melanoma. Its use resulted in 76% of pigmented skin lesions being accurately classified, although these results were based on only five melanomas in a sample of 70 lesions.[4] These preliminary findings were considered encouraging enough to develop the CIA system by improving the software, and modifying the hardware using commercially available items. The aim was to produce a portable CIA system that could be used with ease in a clinical setting. We have described herein the improved results of using the upgraded CIA system as a possible aid in the noninvasive detection of melanoma. In the study of 164 suspected or benign pigmented lesions excised from 129 unselected patients, there were 18 melanomas of which 16 were correctly classified, compared with 15 correctly diagnosed clinically. In contrast, if the clinical ratings of individual characteristics of each lesion (size, color, edge) were used to predict a lesion's diagnostic category, then again 16 of the 18 melanomas and 83% of lesions overall were correctly categorized (compared with 89% with the upgraded CIA system). As previously, the test was intended to be more sensitive than specific in diagnosing melanoma, with the equal prior probabilities that were set in the analyses. However, despite this, the CIA system proved more specific than clinical diagnosis: the CIA system gave 16 false-positive diagnoses of melanoma compared with 22 false-positive clinical diagnoses of melanoma.

The same two melanomas were misdiagnosed by the clinician, by the clinical measures function, and by CIA. They were both of moderate size (5 to 6 mm in diameter) and both were level I. These melanomas represent the typical melanoma that is difficult to diagnose clinically. Herein lies the diagnostic challenge: to distinguish the rare early melanoma from the less serious but common pigmented lesions they so closely resemble. To diagnose the advanced melanoma, clinical aids are hardly necessary, and the potential for curative surgery is small. Although advanced melanomas were not a priori excluded from our study sample, they accounted for few cases; of the 16 melanomas that were correctly classified, about 40% were level III and 50% were level II. Thus our 89% accuracy rate was not inflated by the inclusion of melanomas that were readily diagnosed clinically. Few data are available regarding the diagnostic accuracy that may be achieved by visual inspection alone in relation to early as opposed to advanced melanomas. Cassileth et al.[6] reported that of 105 clinicians (nondermatologists), 30% identified early melanomas as innocuous lesions on the basis of photographs, and of 661 general practitioners in a similar Australian study, 40% misdiagnosed early melanoma.[7] Among 2560 volunteers from the general public in the United States, dermatologists had a positive predictive value of 35% to 40% for all melanomas among lesions that were excised.[8] These comparative findings suggest that CIA, with further development, may indeed be able to assist in the clinical detection of early invasive and noninvasive melanomas.

The features that were the most powerful discriminants of malignant pigmented lesions were first size (large area or perimeter of lesion) and, second, color variegation. Both are well-established clinical features of early melanoma.[9] However, it is the latter feature, the variation of pigmentation within a lesion, that holds the possibility of further refinement with image analysis techniques. This is also the feature that is at the heart of epiluminescence microscopy (ELM), although here diagnosis rests on the examination of epidermal and dermal pigmentary structures that are not visible to the naked eye. The complexity of the ELM grading protocol means that considerable expertise is required for differential diagnosis, although the hope is that advances in imaging technology may facilitate the wider use of digital ELM.[3] Thus ultimately the goals for CIA and ELM are similar and appear achievable, namely, to provide a valuable diagnostic adjunct that will enable clinicians to make highly sensitive and specific diagnostic assessment of the increasingly common,[10] early, curable melanoma.

REFERENCES

1. Witheiler DD, Cockerell CJ. Histologic features and sensitivity of diagnosis of clinically unsuspected cutaneous melanoma. Am J Dermatopathol 1991;13:551-6.
2. Pehamberger H, Binder M, Steiner A, et al. In vivo epiluminescence microscopy: improvement of early diagnosis of melanoma. J Invest Dermatol 1993;100(suppl):356S-62S.
3. Kenet RO, Kang S, Kenet BJ, et al. Clinical diagnosis of pigmented lesions using digital epiluminescence microscopy: grading protocol and atlas. Arch Dermatol 1993; 129:157-74.
4. Green A, Martin N, McKenzie G, et al. Computer image analysis of pigmented skin lesions. Melanoma Res 1991; 1:231-6.

5. Norusis MJ. SPSS/PC+ Advanced Statistics 4.0. SPSS Inc, Chicago, 1990.
6. Cassileth BR, Clark WH Jr, Lusk EJ, et al. How well do physicians recognize melanoma and other problem lesions? J AM ACAD DERMATOL 1986;14:550-60.
7. Paine SL. Early detection of skin cancer: knowledge, practice and beliefs of Victorian general practitioners. MPH thesis, Monash University, 1991.
8. Koh HK, Caruso K, Gage I, et al. Evaluation of melanoma/skin cancer screening in Massachusetts: preliminary results. Cancer 1990;65:375-9.
9. NIH Consensus Conference. Diagnosis and treatment of early melanoma. JAMA 1992;268:1314-8.
10. Burton RC, Coates MS, Hersey P, et al. An analysis of a melanoma epidemic. Int J Cancer 1993;55:765-70.