ELSEVIER

Seminars in Cutaneous Medicine and Surgery

Automated Diagnostic Instruments for Cutaneous Melanoma

Malene E. Vestergaard, MD,* and Scott W. Menzies, MB, BS, PhD†

The objective of this review is to report and discuss the evidence for fully automated diagnostic instruments for cutaneous melanoma tested in a real-world clinical setting directly compared with human diagnosis. A systematic review was performed and articles excluded when studies did not report sensitivity or specificity for melanoma directly compared with humans on an independent test set. Only 3 instruments have had their diagnostic accuracy compared with a human diagnosis in the clinical field with a meaningful sample size that could allow some generalization with the wider clinical arena. Two of these instruments showed a significantly inferior specificity for the diagnosis of melanoma compared with specialists. In one of these studies, the sensitivity for diagnosis, although superior to the specialist diagnosis, did not reach statistical significance. In contrast, one instrument had an equivalent specificity and trended superior but not significantly for sensitivity for the diagnosis of melanoma. Other image based nonclinic studies and studies comparing clinical management as the endpoint rather than diagnosis are also reviewed.

Semin Cutan Med Surg 27:32-36

KEYWORDS melanoma, diagnosis, automated, computer assisted, pigmented lesion, melanocytic

The clinical diagnosis of cutaneous melanoma (CM) is variable and limited depending on the clinician's performance and experience.[1,2] Fully automated diagnostic instruments have been developed with the aim of no or limited input from humans. A systematic review by Rosado and coworkers[3] from 2003 concluded: "The computer diagnosis of melanoma is accurate under experimental conditions but the practical value of automated diagnostic instruments under real world conditions is currently unknown." Today, there are several automated diagnostic instruments available on the market. Ideal requirements for instrument testing have been suggested in the review by Rosado[3] and also recently by Menzies[4] in 2006. These include instrument performance compared with human diagnosis on consecutive or random selected benign and malignant lesions (not only excised lesions because this does not reflect the real world) with clear inclusion/exclusion criteria in a defined clinical setting, tested on an independent set of lesions with repeatability analysis performed and instrument calibration reported. The objective of this review is to report and discuss the evidence for fully automated diagnostic instruments for cutaneous melanoma tested in a real-world clinical setting directly compared with human diagnosis.

Different technologies have been applied with the use of variance in physical properties such as visible and nonvisible light, ultrasound frequencies, magnetic resonance, and electrical impedance in melanoma and nonmelanoma skin lesions. Extensive literature on software with mathematical models for analyzing and classifying the data from skin lesions are published but are beyond the scope of this review. The most thoroughly tested and developed diagnostic systems can be divided into (1) digitized images with diagnostic visual properties automatically segmented sometimes in analogy to traditional dermoscopy and naked eye examination, (2) analysis of sequence of images taken at different wavelengths (multi-spectral images) with similar image analysis to the aforementioned devices or sometimes segmenting

*Department of Dermatology and Sydney Melanoma Diagnostic Centre, Sydney Cancer Centre, Royal Prince Alfred Hospital, Sydney, New South Wales, Australia.

†Sydney Melanoma Diagnostic Centre and Faculty of Medicine, University of Sydney, Sydney Cancer Centre, Royal Prince Alfred Hospital, Sydney, Sydney, New South Wales, Australia.

Supported in part by NSW Melanoma Network, 1A Eden Sreet, North Sydney, NSW 2006, Australia.

Address correspondence to Scott W. Menzies, MB, BS, PhD, Sydney Melanoma Diagnostic Centre, 2nd Floor, Gloucester House, Royal Prince Alfred Hospital, Missenden Road, Camperdown, NSW 2050, Australia. E-mail: scott.menzies@email.cs.nsw.gov.au

1085-5629/08/$-see front matter
doi:10.1016/j.sder.2008.01.001

information of skin chromophores or deeper nonvisualized structures, and (3) electrical impedance devices that uses different electrical properties of malignant skin lesions.

Search Strategy and Selection Criteria

Databases searched were PubMed, MEDLINE, Cochrane, Embase, Clinicial Evidence, and CINAHL from 1987 until 2007. Key words used were Nevus pigmented, diagnostic test, digital imaging, images processing, computer assisted, naked eye examination, skin neoplasm, nevus, melanoma, sensitivity, specificity. This search was provided from The Australian Cancer Network and National Health and Medical Research Council (NHMRC) in conjunction to updating the NHMRC guidelines for melanoma. An additional literature search was performed later with PubMed limited to the period January 1, 2007, to June 1, 2007, searching the keywords melanoma AND computer diagnosis, computer assisted, digital images diagnosis, automated diagnosis, diagnostic test, image processing, neural network, automatic classifier, physical examination, skin neoplasm, pigmented nevus, pigmented lesion. A search on all first and last authors of included studies in the same period was also performed. The titles and abstracts were examined and relevant articles were reviewed by 1 reader (M.V.). Uncertainties were discussed with an expert in the field (S.M.) and resolved by consensus. Articles were excluded when studies did not report sensitivity or specificity for melanoma directly compared with humans on an independent test set. Studies using cross-validation without using an independent test set were not retrieved. Table 1 describes the studies satisfying these requirements. In addition, studies comparing lesion management of the instrument to human performance were retrieved as tabled. When found, studies that did not meet these stringent criteria but were relevant to certain instruments that had satisfied these requirements elsewhere were examined.

Further Developed Devices

We identified 3 instruments that have compared instrument diagnostic accuracy with human diagnosis in a clinical setting using an independent test set and with a sample size large enough to enable generalization to the wider clinical arena. The studies are presented in order of year of publication.

DB-MIPS

Bauer and coworkers[5] tested a diagnostic system on more than 300 patients in a dermatology center in Italy. They used a stereomicroscope and a 3-charge-coupled device (CCD) camera for capturing high-resolution digital dermoscopy images. Image-analysis software, DB-Dermo MIPS® (Dell'Eva-Burroni Dermoscopy Melanoma Image Processing Software, SRL, Siena, Italy) using an artificial neural network (ANN) classifier evaluated image analysis variables of geometries, colors and textures[6] allowing an automated diagnosis of the lesion. The independent test set were 315 consecutive pigmented skin lesions (PSL) preselected for excision (12 in situ CM, 14 invasive CM $<$ 0.75 mm, 10 CM between 0.75 and 1.5 mm, and 6 CM $>$ 1.5 mm Breslow thickness). The human diagnosis was decided by consensus of 3 or 4 dermatologists (1 expert in PSLs) trained in the diagnosis of melanoma and PSLs. The diagnostic system demonstrated higher sensitivity (93%) than the specialists (79%) for diagnosis of melanoma and similar specificity (98% vs 96%). However, this did not reach statistical significance.

DB-MIPS: Other Relevant Studies

In another Italian center, Piccolo and coworkers[7] tested the diagnostic performance of the DEM-MIPS® (Digital Epi Microscopy Melanoma Image Processing Software, Biomips SRL, Siena, Italy) software system using an older release based on a neural network classifier. The independent test set was image based ie, not in the clinical setting, of 13 CM and 328 nonmelanoma lesions, including 281 Clark nevi, excised because of equivocal dermoscopic findings or at the patient's request. They reported 92% sensitivity (95% confidence interval [95% CI] 0.78–1.00) for both the automated system and one trained dermatologist with more than 5 years' experience in dermoscopy. However, the specificity of the system was 74% (95% CI 0.69–0.79), which was significantly lower than the specificity of the expert which reached 99% (95% CI 0.98–1.00). In contrast, a clinician with minimal training in dermoscopy had a sensitivity of 69% (95% CI 0.44–0.94) but a specificity of 94% (95% CI 0.92–0.97) for the diagnosis of melanoma on the same test set.

Finally, use of image analysis features derived from various versions of the DB-Dermo MIPS system to formulate other diagnostic classifiers has been described in a number of settings without comparison with human performance.[8-10]

This automated diagnostic system has been further optimized and developed (DB-MIPS© System, BIO MIPS Engineering, SRL, Siena, Italy). Recently, a large study evaluated the system in 3 clinical centers, 1 referral center, and 2 primary screening centers.[11] The instrument was used as an integrated part of the daily clinical examination and demonstrated a sensitivity of 90% to 95% and specificity 80% to 86% for diagnosis of melanoma depending on the center. No comparison with the specialist clinical diagnosis was reported.

An overall conclusion based on the entire literature of the DB-MIPS® software is difficult because of different classifiers used in studies. In the only study directly comparing the instrument and the specialists in a clinical setting using an older version of DB-MIPS® software, it appeared that the diagnostic performance was at least equivalent to specialist clinicians.[5] However, in the only other study again using an older version of the classifier but also comparing images rather than clinical examination, showed a significantly lower specificity compared with an experienced clinician.[7] Such results may reflect dramatic differences in the proportion of dysplastic nevi in the benign sets examined (Table 1). Further clinical studies of the present instrument are eagerly awaited and will hopefully provide evidence for its diagnostic performance against humans in a defined clinical setting.

Table 1 Studies Comparing Human Diagnosis or Management with Instrument Diagnosis on an Independent Test Set

First Author Year Published	n = Lesions in Test Set	CM	Median Breslow Thickness mm	DN (% of benign lesions)	Lesion Type	Selection of Lesions	Sensitivity Instrument [95% confidence interval]	Sensitivity Clinician [95% confidence interval]	Specificity Instrument [95% confidence interval]	Specificity Clinician [95% confidence interval]
Bauer 2000	315	42	NE	9	PSL	Consecutive to be excised	93	79	98	96
Piccolo 2002	341	13	NR	86	PSL	Consecutive excised	92 [78–100]	92 (TD) [78–100] 69 (MTC) [44–94]	74 [69–79]	99 (TD) [98–100] 94 (MTC) [92–97]
Bono 2002	313	66	0.64	4	PSL	Consecutive to be excised	80	91	49 ($P < 0.001$)	74
Har-Shai 2005	384	53	0.60	27	PSL	Sample to be excised	91	81	58 ($P < 0.001$)	81
Blum 2004	144	32	0.86	53	MSL	Consecutive excised	100 ($P \leq 0.02$) 100 ($P \leq 0.04$) 100 ($P \leq 0.02$)	84 (expert) 88 (average) 84 (beginner)	77 ($P < 0.002$) 77 ($P = 0.022$) 77	92 (expert) 88 (average) 87 (beginner)
Menzies 2005	78	13	NR	66	MSL	Random excised	85 85 85 85	90 (experts) 81 (dermato) 85 (trainee) 62 (GP)	65 65 65 ($P = 0.006$) 65	59 (experts) 60 (dermato) 36 (trainee) 63 (GP)
Carrara 2007	1198	76	NE	NE	PSL	Consecutive patients	NA	NA	NA	NA
Jamora 2003	440	1	In situ	NE	PSL	Atypical	NA	NA	NA	NA
Boldrick 2007	1000	6	0.30	NE	PSL	All PSL on 83 patients	NA	NA	NA	NA

Sensitivity is the number of lesions classified as melanoma divided by the number of lesions with histopathology diagnosis of melanoma, expressed as a percentage. Specificity is the number of lesions classified as non-melanoma divided by the number of lesions with a histopathology diagnosis of non-melanoma, expressed as a percentage. The P-value was calculated by Chi-Square analysis.

CM, Cutaneous Melanoma; DN, Dysplastic Nevus; PSL, Pigmented Skin Lesion; MSL, Melanocytic Skin Lesion; NR, Not Reported; NE, Not Extractable; NA, Not Applicable; TD, Trained Dermatologist; MTC, Minimal Trained Clinician; dermato, Dermatologists; trainee, Dermatologist Trainee; GP, General Practitioner. Expert, average and beginner refer to experience in dermoscopy.

Telespectrometry

Bono and coworkers[12] investigated the performance of a telespectrophotometry (TS) system on 298 patients in Italy in 2002. The system consisted of a CCD camera with 17 filters acquiring spectral images analyzed for reflectance, variegation, pigment distribution, lesion contour and border irregularity parameters and classified by a computer system. The independent test set was 313 consecutive excised PSLs suspicious for melanoma (66 CM, median Breslow thickness 0.64 mm and 247 nonmelanoma including 10 dysplastic nevi [DN]). The clinical diagnosis was performed by 1 of the 4 surgical oncologists with more than 5 years' experience in the diagnosis of PSLs. Their result demonstrated a lower sensitivity of the TS system for melanoma (80%) compared with the sensitivity of the specialists (91%), which failed to reach statistical significance. However, there was a significantly lower specificity for the TS of 49% compared with 74% for the specialists.

Currently another spectroscopy and image analysis-based instrument (MelaFind®, Electro-Optical Sciences, Inc. Irvington, NY) is undergoing a clinical trial with the aim of Food and Drug Administration approval comparing its performance with dermatologists in the United States. The research group of Elbaum and coworkers[13] reported at least 95% sensitivity and close to 70% specificity using cross validation when developing classifiers for the system. The results of this trial are eagerly awaited.

Electrical Impedance

In 2005 Har-Shai and coworkers[14] used an electrical impedance scanning device (TS2000M®, TranScan Medical Ltd., Migdal Ha'Emek, Israel) for the diagnosis of melanoma. The system uses electrodes that measure electrical impedance in the skin lesions. An integrated image analysis system was also developed on a nonindependent set, the results of which are not reported here. The test set of 384 PSLs preselected for excision from trunk or extremities included 53 CM (median Breslow thickness 0.6 mm) and 88 DN. The clinical diagnosis was performed by clinicians at 8 dermatology and surgery centers. The instrument achieved 91% sensitivity for melanoma compared with 81% for the clinician diagnosis, which was not statistically significant. However, the 58% specificity of the instrument was significantly poorer than the 81% specificity for the clinician. The system could not discriminate melanoma on the head or neck due to different impedance properties of the skin at these sites.

Other Instruments That Compare Human Diagnosis in Experimental Conditions or With Smaller Sample Sizes

A dermoscopy image analysis system (Tuebinger Mole Analyzer, University of Tuebingen, Tuebingen, Germany) was tested on an independent set of images (ie, not in a clinical setting) of 144 excised melanocytic skin lesions (MSLs; 32 CM, median Breslow thickness 0.86 mm, 53 melanocytic nevi, and 59 DN) by Blum and coworkers.[15] The automated diagnostic system had a statistically significant higher sensitivity (100%) compared with the 84% sensitivity for both an expert and a beginner in dermoscopy and 88% sensitivity for a clinician with average dermoscopy experience. Specificity of the instrument (77%) was significantly lower compared with the expert (92%) and the average dermoscopy user (88%) and trended but failed to reach statistical significance for the beginner (87%).

The performance of another automated dermoscopy image analysis instrument (SolarScan®, Polartechnics Ltd., Sydney, Australia) was developed and tested by Menzies and coworkers.[16] They simulated the clinical setting by providing 13 clinicians with patient details, lesion history, and clinical and dermoscopic images. The independent test set was 78 MSLs with histopathological diagnosis (13 CM, 22 melanocytic nevi, and 43 DN). The SolarScan® system had a comparable or superior sensitivity and specificity (85% and 65%) compared with those of the mean of 3 international dermoscopy experts (90% and 59%), 4 dermatologists (81% and 60%), 3 dermatology trainees (85% and 36%), and 3 general practitioners (62% and 63%). Only the inferior specificity and positive predictive value for the diagnosis of melanoma for the trainees, and the inferior negative predictive value in general practitioners reached statistically significant differences. Overall, the instrument was reported to have 91% sensitivity and 68% specificity on an independent test set of 122 CM (median Breslow thickness 0.37 mm) and 596 MSLs.

Finally, Barzegari and coworkers[17] also used a diagnostic system based on digital dermoscopy image analysis (microDERM®, VISIOmed AG, Bochum, Germany) on a small sample of 122 PSLs (6 CM and 7 DN) preselected for excision on 91 Iranian patients. While a comparison was made with two clinicians the algorithm threshold was not selected before testing. Nevertheless a threshold was noted that gave the same 83% sensitivity and 96% specificity for both instrument and clinical diagnosis for melanoma.

Instruments Assessed for Management as the Endpoint

Several research groups have explored the performance of automatic diagnostic instruments for selecting PSLs for excision, thus mimicking a specialist clinician management of PSLs. Such an instrument could support nonexpert clinicians to identify PSLs that need investigation.

Carrara and coworkers[18] trained and tested an ANN for classifying PSLs as excision-needing or reassuring based on images acquired by a spectrophotometric image system (SpectroShade®, MHT, Verona, Italy). The automatic classification was compared with the management decision of an expert clinician. In an independent test set of 524 excised PSLs including 76 CM and 674 nonexcised reassuring PSLs the system correctly classified 88% of excised lesions while it only classified 80% of nonexcised lesions as reassuring.

Jamora and coworkers[19] tested whether a commercial available computerized dermoscopy image analyzer (DermoGenius System®, Rodenstock Prazisionsoptik, Munich, Germany) could improve the management of PSLs. One specialist had examined and found 440 PSLs atypical but not sufficient to trigger biopsy. These lesions were analyzed with the instrument and 52 PSLs were biopsied solely on the basis of the system. One in situ melanoma was found in this set by the instrument (ie, missed by clinician) but 51 benign nonmelanoma PSLs were excised (correctly classified by the clinician and incorrectly by the computer system).

This approach was also examined by Boldrick and her research group in an experimental retrospective study.[20] The digital image analyzing system microDERM® (VISIOmed AG, Bochum, Germany) as described above and an expert clinician rated dermoscopy images of 1000 PSLs as benign or worrisome with need for biopsy. The expert rated 18 PSLs worrisome and had complete agreement with the initial clinical assessment to excise. However, the microDERM® system scored 94 PSLs as requiring excision. The "need to biopsy" agreement between the instrument and the expert and the initial clinical examination was poor. Furthermore, the system misclassified 2 of 6 melanoma as not requiring excision among the 18 lesions with a histopathological diagnosis.

Discussion

A generalization of all automated instruments for the diagnosis of melanoma is not appropriate because each offer very different technologies with differing diagnostic abilities. Only 3 instruments have had their diagnostic accuracy compared with a human diagnosis in the clinical field with a meaningful sample size that could allow some generalization with the wider clinical arena. Two of these instruments showed a significantly inferior specificity for the diagnosis of melanoma compared with specialists.[12,14] In one of these studies,[14] the sensitivity for diagnosis, although superior to the specialist diagnosis, did not reach statistical significance possibly because of a deficiency in sample size. In contrast, one instrument[5] had an equivalent specificity and trended superior but not significantly for sensitivity for the diagnosis of melanoma. Again the later result may result from a deficiency in power. Complicating the assessment of this result was the significantly inferior specificity found with the system when comparing an expert diagnosis in an image based study.[7] Finally, the data from these studies is from an older version of the currently available system.

Although the literature on the development and testing of automated diagnostic instruments is extensive, the most exciting question is to assess the proper employment of these instruments in the real world. A crucial issue is the instrument must be tested against human performance to assess the impact of the diagnostic device in the clinical arena. This will allow assessment of the nonmelanoma set (specificity) which may vary dramatically as the level of clinical atypia of these benign lesions varies from study to study. As well, assessment of an instruments ability to detect clinically difficult melanoma is perhaps best understood when comparing performance against humans. Furthermore, studies including lesions not preselected for excision with follow up to ensure the true benign diagnosis would allow better assessment of the impact in the field.

References

1. Kittler H, Pehamberger H, Wolff K, et al: Diagnostic accuracy of dermoscopy. Lancet Oncol 3:159-165, 2002
2. Bafounta M, Beauchet A, Aegerter P, et al: Is dermoscopy (epiluminescence microscopy) useful for the diagnosis of melanoma? Results of a meta-analysis using techniques adapted to the evaluation of diagnostic tests. Arch Dermatol 137:1343-1350, 2001
3. Rosado B, Menzies S, Harbauer A, et al: Accuracy of computer diagnosis of melanoma: A quantitative meta-analysis. Arch Dermatol 139:361-367, 2003
4. Menzies SW: Cutaneous melanoma: Making a clinical diagnosis, present and future. Dermatol Ther 19:32-39, 2006
5. 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 10:345-349, 2000
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 135:1459-1465, 1999
7. Piccolo D, Ferrari A, Peris K, et al: 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. Br J Dermatol 147:481-486, 2002
8. Seidenari S, Pellacani G, Giannetti A: Digital videomicroscopy and image analysis with automatic classification for detection of thin melanomas. Melanoma Res 9:163-171, 1999
9. Burroni M, Corona R, Dell'Eva G, et al: Melanoma computer-aided diagnosis: Reliability and feasibility study. Clin Cancer Res 10:1881-1886, 2004
10. Burroni M, Sbano P, Cevenini G, et al: Dysplastic naevus vs. in situ melanoma: Digital dermoscopy analysis. Br J Dermatol 152:679-684, 2005
11. Wollina U, Burroni M, Torricelli R, et al: Digital dermoscopy in clinical practise: A three-centre analysis. Skin Res Technol 13:133-142, 2007
12. Bono A, Bartoli C, Cascinelli N, et al: Melanoma detection. A prospective study comparing diagnosis with the naked eye, dermoscopy and telespectrophotomatry. Dermatology 205:362-366, 2002
13. Elbaum M, Kopf AW, Rabinovitz HS, et al: Automatic differentiation of melanoma from melanocytic nevi with multispectral digital dermoscopy: A feasibility study. J Am Acad Dermatol 44:207-218, 2001
14. Har-Shai Y, Glickman YA, Siller G, et al: Electrical impedance scanning for melanoma diagnosis: a validation study. Plast Reconstr Surg 116: 782-790, 2005
15. Blum A, Hofmann-Wellenhof R, Luedtke H, et al: Value of the clinical history for different users of dermoscopy compared with results of digital image analysis. J Eur Acad Dermatol Venereol 18:665-669, 2004
16. Menzies SW, Bischof L, Talbot H, et al: The performance of SolarScan: An automated dermoscopy image analysis instrument for the diagnosis of primary melanoma. Arch Dermatol 141:1388-1396, 2005
17. Barzegari M, Ghaninezhad H, Mansoori P, et al: Computer-aided dermoscopy for diagnosis of melanoma. BMC Dermatol 65:8, 2005
18. Carrara M, Bono A, Bartoli C, et al: Multispectral imaging and artificial neural network: Mimicking the management decision of the clinician facing pigmented skin lesions. Phys Med Biol 52:2599-2613, 2007
19. Jamora MJ, Wainwright BD, Meehan SA, et al: Improved identification of potentially dangerous pigmented skin lesions by computerized image analysis. Arch Dermatol 139:195-198, 2003
20. Boldrick J, Layton C, Nguyen J, et al: Evaluation of digital dermoscopy in a pigmented lesion clinic: Clinician versus computer assessment of malignancy risk. J Am Acad Dermatol 56:417-421, 2007