BMJ Open

The feasibility of using manual segmentation in a multifeature computer-aided diagnosis system for classification of skin lesions: a retrospective comparative study

Wen-Yu Chang,[1,2] Adam Huang,[3] Yin-Chun Chen,[4,5] Chi-Wei Lin,[6,7] John Tsai,[5,8] Chung-Kai Yang,[9] Yin-Tseng Huang,[10] Yi-Fan Wu,[11,12] Gwo-Shing Chen[13]

To cite: Chang W-Y, Huang A, Chen Y-C, *et al*. The feasibility of using manual segmentation in a multifeature computer-aided diagnosis system for classification of skin lesions: a retrospective comparative study. *BMJ Open* 2015;**5**:e007823. doi:10.1136/bmjopen-2015-007823

▸ Prepublication history for this paper is available online. To view these files please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2015-007823).

Received 29 January 2015
Revised 27 March 2015
Accepted 8 April 2015

For numbered affiliations see end of article.

Correspondence to
Professor Gwo-Shing Chen; d700086@kmu.edu.tw

ABSTRACT

Objectives: To investigate the feasibility of manual segmentation by users of different backgrounds in a previously developed multifeature computer-aided diagnosis (CADx) system to classify melanocytic and non-melanocytic skin lesions based on conventional digital photographic images.

Methods: In total, 347 conventional photographs of melanocytic and non-melanocytic skin lesions were retrospectively reviewed, and manually segmented by two groups of physicians, dermatologists and general practitioners, as well as by an automated segmentation software program, JSEG. The performance of CADx based on inputs from these two groups of physicians and that of the JSEG program was compared using feature agreement analysis.

Results: The estimated area under the receiver operating characteristic curve for classification of benign or malignant skin lesions based were comparable on individual segmentation by the gold standard (0.893, 95% CI 0.856 to 0.930), dermatologists (0.886, 95% CI 0.863 to 0.908), general practitioners (0.883, 95% CI 0.864 to 0.903) and JSEG (0.856, 95% CI 0.812 to 0.899). The agreement in the malignancy probability scores among the physicians was excellent (intraclass correlation coefficient: 0.91). By selecting an optimal cut-off value of malignancy probability score, the sensitivity and specificity were 80.07% and 81.47% for dermatologists and 79.90% and 80.20% for general practitioners.

Conclusions: This study suggests that manual segmentation by general practitioners is feasible in the described CADx system for classifying benign and malignant skin lesions.

Strengths and limitations of this study

- SKINCAD, an effective computer-aided diagnosis (CADx) system developed in our previous study, achieved performance similar to face-to-face clinical diagnosis by staff dermatologists at our institution. Regarding clinical application, evaluating the feasibility of subjective manual segmentation by users of different backgrounds is especially useful for images with complicated components, whereas automated segmentation methods sometimes fail.
- This study simulates a real clinical setting and groups, general practitioners (GPs) and dermatologists, have manually segmented a wide spectrum of skin lesions under generous inclusion criteria, which represents skin lesions encountered in daily practice, with each lesion given a definite histopathological diagnosis.
- Our result suggests that GP-determined borders performed as well as dermatologist-determined borders in CADx diagnosis by SKINCAD for classification of benign or malignant skin lesions.
- Through the in-depth evaluation of overlap index and feature agreement levels, our study indicates the possibility of direct onsite computation application for physicians other than dermatology specialists when assessing melanocytic and non-melanocytic skin lesions.
- This retrospective analysis was restricted to biopsied lesions performed in a single medical centre, as we used the histopathological reports as the gold standard. Further large-scale prospective study may be required in the future for broader application.

INTRODUCTION

Skin cancer is a common malignancy worldwide. The increasing cost of skin cancer management over the last decade constitutes a substantial health problem.[1–3] With lower incidence rates of melanoma in Asians than in Caucasians, non-melanoma skin cancers, such as squamous cell carcinoma and basal cell carcinoma, contribute to significant morbidities as well, especially in the Asian

population. In clinical practice, most physicians detect skin cancer by visual examination, which is highly dependent on experience and specialisation. Although the accuracy rate of clinicians can be improved with the support of dermoscopy when confronted with difficult-to-diagnose skin lesions, this approach relies on the specific training of a limited population of clinicians, mainly dermatological specialists who manage skin tumours.

Previously, we developed an effective computer-aided diagnosis (CADx) system (SKINCAD), which classifies melanocytic and non-melanocytic skin lesions by utilising conventional digital macrophotographs. This system achieved performance similar to face-to-face clinical diagnosis by staff dermatologists at our institution.[4] In the study, a dermatologist manually segmented the images for analysis. Regarding the concerns of subjectivity and consistency of the manually generated borders,[5 6] many automatic border detection methods have been developed, such as the JSEG algorithm, contrast enhancement and clustering algorithms, with different evaluation metrics.[6–11] However, many of these algorithms were developed to approximate the ground truth borders, which were also determined by dermatologists subjectively.[7 8 10 12 13] Furthermore, because of the complexity of skin images, it is not usually possible for every lesion to be segmented automatically. In a study that compared three dermoscopic image analysis systems, approximately half of the skin lesions were not analysable by at least one of the three systems due to programming limitations, such as the inability to perform segmentation, and the operator had to adjust the computer-determined segmentation manually.[14] We would expect that the segmentation task might be more challenging when clinical digital photographs are used.

In our previous study, the use of the CADx system could be repeated with relative consistency by users without medical training.[15] The purpose of this study was to investigate the feasibility and reliability of manual segmentation performed by medical practitioners of different backgrounds (general practitioners (GPs) or board-certified dermatologists), and to compare their performance with that of a commonly used autosegmentation algorithm, JSEG, in a multifeature, CADx system, SKINCAD. In particular, this study aimed to assess the potential use of this system by GPs without skin cancer training.

MATERIALS AND METHODS

Data acquisition

From January 2010 to December 2010, 2148 consecutive skin lesions were biopsied or excised by dermatologists for histological confirmation in the Department of Dermatology, Kaohsiung Medical University. A total of 13 908 digital photographs of these lesions were taken, prior to the biopsy and excision procedures, for recording purposes. We retrospectively reviewed all cases and excluded non-tumour specimens, lesions that had undergone previous surgical procedures and images that were misregistered or of poor quality (unfocused or motion blurred). The database consists of 1151 images from 295 patients (347 specimens). All of the lesions were examined at the clinic by board-certified staff dermatologists from our institute.

A dermatologist (W-YC, with 11 years of experience) carefully reviewed these images and selected a representative close-up image for each lesion. Images with hair artefacts were not excluded because the preselection of data showed little influence based on a large data set in a previous report.[16] Photographs were obtained with a 6.1-megapixel digital single-lens reflex camera (D70, Nikon Corporation, Tokyo, Japan) with an 18–50 mm F2.8 macrolens (Sigma Corporation, Fukushima, Japan). When obtaining close-up images, the target lesion was focused and positioned at the centre of the photographs, and the size was controlled so as not to exceed 50% of the area of the photograph.

CADx system

The in-house CADx system used in this study, SKINCAD, was developed and optimised in our previous report to include 16 effective features, with optimal operation parameters for an Asian database.[4] In brief, it is a CADx system that applies various image processing algorithms to extract the shape, colour and texture features, not only for the skin lesion itself but also for the surrounding rectangular cropped area. The raw score assigned after the CADx evaluation is adjusted to be the probability of malignancy according to the previous database model calibrated using Platt's method.[4 17]

Manual segmentation study

Six months after the images were collected, four board-certified family physicians and three dermatologists (including W-YC) with an average experience of 10 years were asked to draw the skin lesion borders for all 347 digital images using the CADx system's graphical user interface (GUI), which was developed using MATLAB software (figure 1). When the physician started to use the software, a brief introduction was presented with 10 sample images demonstrating how to use the software to mark the borders and how to save the results. The software GUI showed one lesion at a time, and the reader could use the pan/zoom function as needed for observation. To simulate a real clinical setting, there were no preselection processes, specific instructions or feedback regarding the actual performance of their segmentation results. The physicians evaluated and marked the lesion border according to their own clinical experience, and they were allowed to use their own equipment for convenience, as long as they were confident enough to accurately define the tumour lesion area and differentiate it from normal skin. Two physicians chose to use a pen tablet (CTH-661, Wacom, Taiwan), and the others

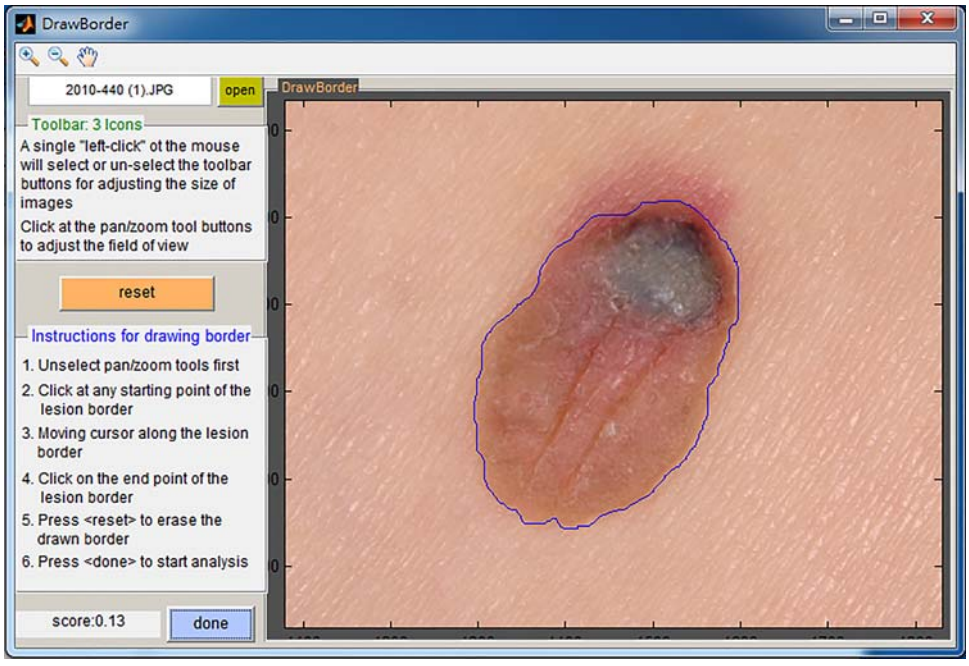

Figure 1 The SKINCAD system graphical user interface developed using MATLAB software.

chose to use their own mouse as a pointing and border-drawing device.

Automated segmentation study

There are many image segmentation algorithms, and it is challenging to compare them. In our previous report, we evaluated a few software programs that are available online,[8 18 19] and tested them by using a diverse data set of 769 images with 19 categories of histological diagnoses.[15] Owing to the complexity of the clinical images, a non-medical individual performed the preselection of an area of interest prior to the automated segmentation phase of all methods. Among the methods we tested preliminarily, JSEG algorithm[18] performed the best with respect to all three criteria, those being, good colour capability, short computational time and consistent accuracy. JSEG was chosen as our automated segmentation algorithm because of its flexibility and good performance in a variety of domains such as natural scenery,[18] colonoscopy images,[20] tongue images[21] and skin lesion images.[13]

Overlapping and variability

The average or comparable extraction results obtained by multiple dermatologists can be considered more reliable.[12] In this study, we defined the area extracted by two or more dermatologists as the gold standard of the ground-truth tumour area. To assess the reliability of the performance of CADx using manual segmentation and the variability introduced by each user, we computed the following:

1. Intraclass correlation coefficient (ICC) to assess inter-rater reliability of the malignancy probability scores among multiple raters.[22]
2. Area overlap percentage or Jaccard index,[23] as defined in Eq. (1), between each GP and the gold standard

$$\text{Jaccard index} = \frac{S_{GP} \cap S_{\text{Gold standard}}}{S_{GP} \cup S_{\text{Gold standard}}} \quad (1)$$

where S_{GP} and $S_{\text{Gold standard}}$ denote the lesion areas determined by the borders drawn by GP and the gold standard, respectively, and and $S_{GP} \cup S_{\text{Gold standard}}$ are the intersection and union, respectively.

3. Lesion inclusion rate: the percentage of the ground-truth lesion included in the cropped image area (rectangular image generated by software to enclose the entire manual segmentation) as defined in Eq. (2)

$$\text{Lesion inclusion rate} = \frac{S_{\text{Gold standard}} \cap S_{\text{Cropped image}}}{S_{\text{Gold standard}}} \quad (2)$$

The above definitions are illustrated in figure 2.

Statistics

The performance of CADx was evaluated based on receiver operating characteristic (ROC) curve analysis, with the pathological results considered as the gold standard of malignancy diagnosis. The area under the ROC curve (Az) was estimated after each physician using CADx. The Az values of two ROC curves were compared using DeLong's test.[24] The masks generated from skin lesion segmentation by all seven physicians and JSEG were compared based on the Jaccard coefficient.[23] The Wilcoxon rank sum test was used to compare the calculated Jaccard indices between the GPs and JSEG. Separate ICCs were computed to assess the contribution

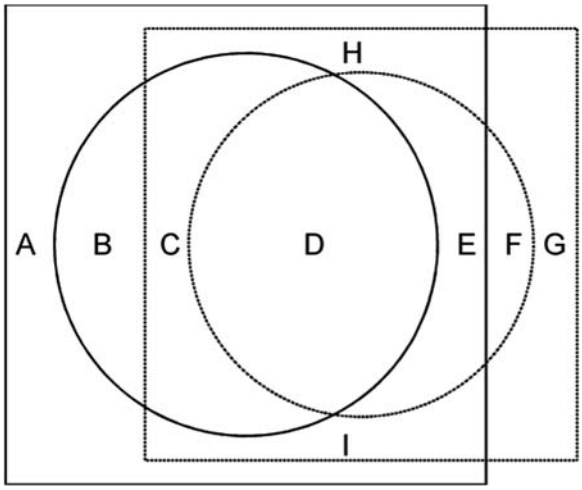

Figure 2 In this illustration, the lesions defined by the gold standard (circle by solid line) and general practitioner (GP; circle by dotted line) are areas B+C+D and D+E+F, respectively. The cropped image generated by GP (rectangle by dotted line) is presented as area C+D+E+F+G+H+I. The Jaccard index by definition is the area overlap percentage defined between by each GP and the gold standard, which is D/(B+C+D+E+F). The lesion inclusion rate is the percentage of ground-truth area included in the cropped image area generated by GP, that is, (C+D)/(B+C+D) in this illustration. With a high lesion inclusion rate, the main differences between each rectangular lesion derived from the marked borders by each GP and the gold standard may primarily involve background peripheral normal skin (F+G and A), which could be assumed to have similar characteristics.

of different lesion borders determined by different readers to the 16 feature scores and the probability score. The ICC was calculated using a two-way random model with measures of absolute agreement.[22] All statistical analyses were performed using R V.3.1.0.[25]

RESULTS

Demographics

From January 2010 to December 2010, a total of 347 images of distinct regions of interest were obtained from 295 patients, including 124 males (42%) and 171 females (58%) with a mean age of 50.9±20.4 years. These images included 97 malignant lesions and 250 benign lesions. The demographic data associated with each histological diagnosis are summarised in table 1. There were seven invasive melanomas in this study. The number of melanomas in our database was small, consistent with the relatively low incidence in the Asian population compared with Caucasians.

CADx performance

To compare the performance of CADx used by the dermatologists and GPs, the Az values based on individual manual segmentation by three dermatologists and four GPs were calculated and compared (figure 3). The border determination performance based on the gold

Table 1 Demographic data for each histological diagnosis

Pathology	N	Per cent	Sex (F/M)	Mean age (year)
Benign	250	72.05	156/93	43.69
Blue naevus	12	3.46	9/3	42.67
Compound naevus	25	7.20	18/6	32.52
Dermatofibroma	3	0.86	1/2	35.67
Haemangioma	13	3.75	7/6	53.00
Intradermal naevus	109	31.41	72/37	37.84
Junctional naevus	21	6.05	14/7	33.10
Seborrheic keratosis	67	19.31	35/32	59.42
Malignant	97	27.95	43/54	69.75
Basal cell carcinoma	47	13.54	23/24	71.17
Bowen's disease	17	4.90	8/9	68.29
Kaposi's sarcoma	4	1.15	0/4	48.25
Keratoacanthoma	3	0.86	1/2	56.00
Melanoma	7	2.02	5/2	63.86
Squamous cell carcinoma	19	5.48	6/13	76.42
All lesions	347	100.00	199/147	50.97

standard was also assessed as previously described. The Az values of the CADx system for classification of benign or malignant skin lesions were as follows: 0.893 (95% CI 0.856 to 0.930) when segmented by the gold standard, 0.886 (95% CI 0.863 to 0.908) when segmented by the

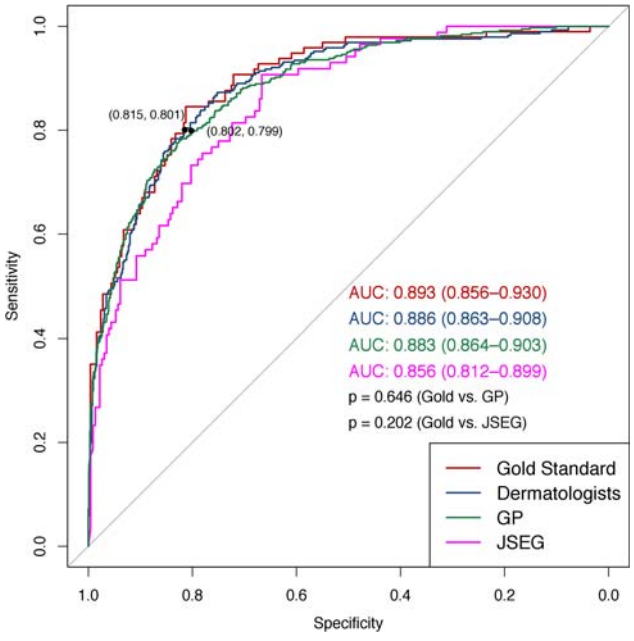

Figure 3 The performance of the discrimination of skin malignancy by four ROC curves generated from segmentation results produced by the gold standard, dermatologists, GP and JSEG. Please note that the 34 failure cases of autosegmentation (9.8%) by JSEG are not included in the ROC analysis (AUC, area under curve; GP, general practitioner; ROC, receiver operating characteristic).

dermatologists, 0.883 (95% CI 0.864 to 0.903) when segmented by the GPs and 0.856 (95% CI 0.812 to 0.899) when segmented by JSEG. The Az values of the dermatologists were slightly better than those of the GPs, but there were no significant differences between the performances generated by the gold standard and the GPs (p=0.65). By adjusting the SKINCAD software operating point based on an optimal cut-off value by seven physicians at 0.3183, the sensitivity and specificity were 80.07% and 81.47%, respectively, for the dermatologists, which were comparable to the values of 79.90% and 80.12%, respectively, for the GPs. All physicians were able to complete the manual segmentation of each lesion in less than 1 min and the probability score could be generated by the CADx within 30 s. The average JSEG computation time is 8.9±8.1 s for each lesion. Owing to the complexity of the clinical images, an extra 10–15 s was required prior to the JSEG segmentation process, to manually preselect an area of interest.

Overlapping results

After omitting 34 (9.8%, 34/347) cases in which JSEG failed at border detection, 313 (90.2%) gold standard images were used to compare the JSEG and GP results (table 2). As the gold standard for overlap evaluation was derived from the dermatologists' original markings, results of the dermatologists were not included for analysis.

The overall Jaccard index of the lesions segmented by the GPs and JSEG compared with the gold standard was 0.70±0.15 and 0.60±0.27, respectively. The GPs performed better than JSEG with respect to the Jaccard index in the benign and malignant categories (p<0.01). When using cropped images for lesion inclusion rate (the percentage of the ground-truth lesion included in the cropped image area derived from user-defined borders), the overlap result using borders derived from GPs reached a higher inclusion ratio (0.96±0.10) compared with that derived from JSEG (0.85±0.31) for benign and malignant lesions (p<0.01).

Interobserver variability

The agreement associated with 16 individual features using different segmented areas generated by the seven physicians is summarised in table 3. These features are listed in order according to the recursive feature elimination ranking in our previous study.[4] They are the foundation of CADx computation, and high agreement indicates CADx scoring consistency. The agreement in compactness and radial variance between the seven physicians was 0.57 and 0.63, respectively (fair-to-good level). The agreement regarding texture features, including grey level run length matrix and coarseness, was excellent (0.84–0.94). Colour features, for example, PC3, showed excellent agreement at a level of 0.96–0.98, and the conventional colour features also reached an excellent agreement level of 0.79–0.97 among all the physicians. The overall probability score derived from the seven physicians reached excellent agreement (0.91). We also investigated the agreement in feature scores between the individual GPs and the gold standard derived from dermatologists. All GPs reached excellent agreement in the final probability score and all 14 features (0.77–0.99), with the exception of compactness and radial variance (0.52–0.75), whereas there was only fair agreement (0.48) between JSEG and the gold standard in the final probability score, and poor agreement in the compactness and radial variance features at 0.13 and 0.29, respectively.

Table 2 The Jaccard index and lesion inclusion rate of GP and JSEG, compared with the gold standard

	Jaccard index			Lesion inclusion rate		
Pathology	**GP**	**JSEG**	**p Value**	**GP**	**JSEG**	**p Value**
Benign	0.68±0.15	0.60±0.26	0.00	0.96±0.10	0.86±0.30	0.00
Blue naevus	0.64±0.14	0.65±0.10	0.88	0.94±0.11	0.95±0.05	0.60
Compound naevus	0.70±0.14	0.62±0.23	0.10	0.96±0.09	0.92±0.21	0.12
Dermatofibroma	0.68±0.13	0.81±0.10	0.20	0.96±0.05	1.00±0.00	0.15
Haemangioma	0.65±0.20	0.63±0.24	0.82	0.99±0.03	0.96±0.10	0.32
Junctional naevus	0.61±0.16	0.58±0.27	0.02	0.91±0.15	0.83±0.34	0.00
Intradermal naevus	0.68±0.14	0.51±0.29	0.37	0.95±0.10	0.74±0.39	0.04
Seborrheic keratosis	0.72±0.13	0.62±0.28	0.07	0.97±0.11	0.87±0.28	0.00
Malignant	0.73±0.15	0.60±0.28	0.00	0.98±0.08	0.84±0.32	0.00
Basal cell carcinoma	0.72±0.14	0.65±0.25	0.21	0.97±0.08	0.88±0.28	0.06
Bowen's disease	0.79±0.10	0.56±0.36	0.03	1.00±0.01	0.77±0.39	0.00
Kaposi's sarcoma	0.75±0.11	0.76±0.10	0.89	0.98±0.07	1.00±0.00	0.08
Keratoacanthoma	0.66±0.17	0.51±0.45	0.95	0.94±0.14	0.67±0.58	0.82
Melanoma	0.75±0.12	0.57±0.29	0.10	0.98±0.03	0.88±0.21	0.31
Squamous cell carcinoma	0.68±0.19	0.50±0.30	0.01	0.98±0.12	0.78±0.37	0.00
All lesions	0.70±0.15	0.60±0.27	0.00	0.96±0.10	0.85±0.31	0.00

Note that 34 cases in which JSEG failed for border detection were not included in the analysis.
GP, general practitioner.

Table 3 The agreement scores of each feature generated from segmentation results by all users, JSEG and the gold standard, assessed by ICC

Features	All 7 physicians	GP 1 vs gold	GP 2 vs gold	GP 3 vs gold	GP 4 vs gold	JSEG vs gold
PC3*	0.96	0.97	0.97	0.99	0.98	0.95
Variance blue channel*	0.88	0.89	0.90	0.95	0.92	0.84
Variance blue channel†	0.97	0.97	0.97	0.99	0.98	0.98
Compactness*	0.57	0.52	0.57	0.70	0.64	0.13
Radial variance*	0.63	0.67	0.60	0.75	0.58	0.29
Green–blue correlation*	0.77	0.87	0.77	0.80	0.87	0.65
Green–grey correlation*	0.80	0.88	0.82	0.88	0.87	0.66
PC3†	0.98	0.98	0.97	0.99	0.99	0.95
Entropy red channel†	0.90	0.94	0.89	0.95	0.90	0.93
Entropy red channel*	0.92	0.96	0.90	0.97	0.94	0.86
Entropy blue channel†	0.93	0.93	0.94	0.97	0.96	0.94
Entropy blue channel*	0.88	0.93	0.88	0.94	0.90	0.82
GLRLM_HGRE_4Level†	0.84	0.91	0.77	0.93	0.87	0.86
GLRLM_SRLGE_4Level†	0.95	0.96	0.95	0.98	0.97	0.93
GLRLM_SRLGE_2Level†	0.94	0.94	0.94	0.97	0.95	0.94
Tamura's coarseness features*	0.94	0.93	0.93	0.97	0.95	0.91
Probability score	0.91	0.91	0.90	0.94	0.93	0.48

Features are listed in order according to RFE ranking between 91 features in our previous study. All of the p values of each ICCs in this table are ≤0.01. The failure cases (34/347) of autosegmentation by JSEG are not included in the analysis.
*Derived from the lesion area only.
†Derived from the whole cropped image.
GLRLM, grey level run length matrix; GP, general practitioner; HGRE, high grey level run emphasis; ICC, intraclass correlation coefficient; PC3, the variance along the coordinates of the third principal components; RFE, recursive feature elimination; SRLGE, short run low grey level emphasis.

DISCUSSION

Precise segmentation was considered an essential step when using the CADx system for skin cancer diagnosis.[4] Without the use of microscopic facilities, there were no definite criteria for defining the true borders between the lesion and non-lesion areas under gross examination, even by expert dermatologists.[26] Therefore, subjectivity is always a concern, and automated segmentation may not always be applicable to these clinical situations. There remained no unified results available for comparing all the tested segmentation algorithms due to differences in the ground-truth definitions and evaluation metrics.[27–29] When utilising conventional digital photographs for the analysis of melanocytic and non-melanocytic lesions, the images usually consist of more colours than the dermoscopic images of melanocytic lesions. Regarding clinical application, subjective manual segmentation in SKINCAD is especially useful for images with complicated components, as human eyes are good at pattern recognition for diagnosis, whereas automated segmentation methods sometimes fail. The failure rate of JSEG was 9.8% (34/347) in our study, similar to the situation of a previous report based on dermoscopic images in real clinical settings in which cases may be rejected for analysis by CADx due to unsuccessful autosegmentation.[14] In addition, our software provides an easy-to-use interface for manual segmentation. Users are able to complete the process in less than 2 min for each lesion, including the manual segmentation process and the probability score computation.

The strength of this study lies in the fact that the groups of GPs and dermatologists have both manually segmented a wide spectrum of skin lesions under generous inclusion criteria, which represents skin lesions encountered in daily practice, with each lesion given a definite histopathological diagnosis. SKINCAD performed as well as face-to-face clinical diagnosis by staff dermatologists at our institution in our previous report.[4] In this study, with a new data set unknown to all raters and SKINCAD, SKINCAD achieved good Az performance for classification of benign or malignant skin lesions using borders determined by either GPs, dermatologists or the gold standard developed from dermatologists' original markings. Instead of testing segmentation methods to merely approximate the ground-truth tumour area determined by dermatologists, which may be of concern as standard, we successfully validated the system's reliability with melanocytic and non-melanocytic skin lesion classification based on a good-sized data set with real clinical settings by assessing the robustness with respect to each feature and final probability score agreement levels derived from borders segmented by individual users of different backgrounds.

We introduced the agreement in 16 features as a performance metric. SKINCAD performed well with colour features showing excellent agreement (0.77–0.98) between the two groups of physicians. The segmentation differences between the users resulted in discrepancies mainly with respect to the border-derived features of compactness and radial variance; only fair-to-good agreement

was reached for these features. Given the known individual subjectivity of each physician, and the results of agreement with the dermatologist-derived gold standard regarding all features and the final probability score, GPs reached a stable performance that was better than that achieved by JSEG. Generally, human users performed better than JSEG in all overlap indices. As 7 of 14 colour and texture features were derived from the cropped rectangular images generated from each crooked border determined by the users, the images for analysis consisted of the main tumour lesion area and peripheral background skin during the preprocessing of the peripheral extension. Therefore, the lesion inclusion rate, which was 0.96±0.1 in this study, also contributed to the stability of the analysis results. This result indicates that a very high proportion of the main lesion on average was included in the cropped rectangular images used for analysis in spite of the discrepancies in the borders drawn by each GP. With a high lesion inclusion rate, the main differences between each rectangular lesion derived from the marked borders by each GP and the gold standard may primarily involve background peripheral normal skin, which could be assumed to have similar characteristics (figure 2). This result may also explain that despite an average Jaccard index of 0.70 ±0.15 between the GPs and the gold standard, the evaluations by all of the physicians still achieved excellent agreement with respect to most features and the final probability scores. A consistent performance in colour features was maintained with subjective manual segmented border variation. This implies that when other useful features, besides shape features, were selected in the SKINCAD, the segmentation variation related to borders may have impacted less on the classification accuracy than a system that uses border-sensitive features only.

There were limitations in this study. All images were obtained from patients visiting a single centre in southern Taiwan using a single image-capture system with consistent quality control for each photograph. The analysis was retrospective and restricted to biopsied lesions, as we used the histopathological reports as the gold standard. We were unable to evaluate the performance regarding lesions for which clinicians or patients decided not to perform the biopsy, as they were not included in the data set. The performance of this system was not compared with that in other CADx studies due to different clinical settings. Further large-scale prospective study may be required in the future for broader application.

In conclusion, our study established a possible model for the diagnosis of skin lesions using conventional digital photography with manual segmentation. A system with multiple features, including border-sensitive and non-border-sensitive features, may compensate for the impact of the subjectivity of manual segmentation. This may be an appropriate solution especially when automatic segmentation is not feasible or applicable. Through the in-depth evaluation of overlap index and feature agreement levels, our study indicates the possibility of direct onsite computation application for physicians other than dermatology specialists, when assessing skin lesions. By combining effective feature extractions by modern computation technologies, and manual segmentations of the lesion area and peripheral skin, SKINCAD may play a role as a consistent second opinion for dermatologists and for GPs. Research on the benefits of SKINCAD with respect to clinical decision-making improvement should be performed in the future.

Author affiliations
[1]Graduate Institute of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan
[2]Department of Dermatology, E-Da Hospital, I-Shou University, Kaohsiung, Taiwan
[3]Research Center for Adaptive Data Analysis, Center for Biomarkers and Translational Medicine, and Graduate Institute of Biomedical Engineering, National Central University, Jhongli, Taiwan
[4]Department of Dermatology, Kaohsiung Municipal Ta-Tung Hospital, Kaohsiung, Taiwan
[5]Department of Dermatology, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan
[6]Department of Family Medicine, E-Da Hospital, I-Shou University, Kaohsiung, Taiwan
[7]Graduate Institute of Adult Education, National Kaohsiung Normal University, Kaohsiung, Taiwan
[8]Chiu Kun-Hsing Dermatology Clinic, Kaohsiung, Taiwan
[9]Huang Yung-Hsueh Clinic, Changhua, Taiwan
[10]Health Management Center, Far Eastern Memorial Hospital, New Taipei City, Taiwan
[11]Department of Family Medicine, Taipei City Hospital, Renai Branch, Taipei, Taiwan
[12]Department of Psychology, National Chengchi University, Taipei, Taiwan
[13]Department of Dermatology, Kaohsiung Medical University Hospital. Department of Dermatology, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan

Acknowledgements The authors would like to thank Professor Chien-Hung Lee and Dr Hui-Min Hsieh for their advice on statistical analysis.

Contributors W-YC and AH were responsible for the overall design, development and refinement of the protocol. G-SC led the research, and supervised on the design and conduct of the analysis. W-YC, AH, Y-CC, C-WL, JT, C-KY, Y-TH and Y-FW were involved in acquisition and interpretation of the data, and the development of the initial protocol. AH provided methodological expertise. W-YC and AH performed statistical analysis. Y-CC and C-WL assisted with the development and refinement of the protocol. All authors made substantial contributions to the drafting, critical revision of important intellectual content and final approval of the document.

Funding This research was supported by the Ministry of Science and Technology of Republic of China (Taiwan) under Grants NSC 100-2221-E-008-009, 101-2221-E-008-021 and MOST 103-2911-I-008-001.

Competing interests None declared.

Ethics approval This study was approved by the institutional review board of the Kaohsiung Medical University Hospital (KMUH-IRB-20130236).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The study database consists of skin lesion images obtained for clinical records. This database cannot be shared without approval from the institutional review board of the Kaohsiung Medical University Hospital.

REFERENCES

1. Rogers HW, Coldiron BM. Analysis of skin cancer treatment and costs in the United States Medicare population, 1996–2008. *Dermatol Surg* 2013;39:35–42.
2. Rogers HW, Weinstock MA, Harris AR, *et al*. Incidence estimate of nonmelanoma skin cancer in the United States, 2006. *Arch Dermatol* 2010;146:283–7.
3. Housman TS, Feldman SR, Williford PM, *et al*. Skin cancer is among the most costly of all cancers to treat for the Medicare population. *J Am Acad Dermatol* 2003;48:425–9.
4. Chang WY, Huang A, Yang CY, *et al*. Computer-aided diagnosis of skin lesions using conventional digital photography: a reliability and feasibility study. *PLoS ONE* 2013;8:e76212.
5. Schmid-Saugeona P, Guillodb J, Thirana JP. Towards a computer-aided diagnosis system for pigmented skin lesions. *Comput Med Imaging Graph* 2003;27:65–78.
6. Joel G, Philippe SS, David G, *et al*. Validation of segmentation techniques for digital dermoscopy. *Skin Res Technol* 2002;8:240–9.
7. Abbas Q, Garcia IF, Emre Celebi M, *et al*. Unified approach for lesion border detection based on mixture modeling and local entropy thresholding. *Skin Res Technol* 2013;19:314–19.
8. Abbas Q, Garcia IF, Emre Celebi M, *et al*. A perceptually oriented method for contrast enhancement and segmentation of dermoscopy images. *Skin Res Technol* 2013;19:e490–7.
9. Suer S, Kockara S, Mete M. An improved border detection in dermoscopy images for density based clustering. *BMC Bioinformatics* 2011;12(Suppl 10):S12.
10. Garnavi R, Aldeen M, Celebi ME, *et al*. Border detection in dermoscopy images using hybrid thresholding on optimized color channels. *Comput Med Imaging Graph* 2011;35:105–15.
11. Celebi ME, Iyatomi H, Schaefer G, *et al*. Lesion border detection in dermoscopy images. *Comput Med Imaging Graph* 2009;33:148–53.
12. Iyatomi H, Oka H, Saito M, *et al*. Quantitative assessment of tumour extraction from dermoscopy images and evaluation of computer-based extraction methods for an automatic melanoma diagnostic system. *Melanoma Res* 2006;16:183–90.
13. Emre Celebi M, Alp Aslandogan Y, Stoecker WV, *et al*. Unsupervised border detection in dermoscopy images. *Skin Res Technol* 2007;13:454–62.
14. Perrinaud A, Gaide O, French L, *et al*. Can automated dermoscopy image analysis instruments provide added benefit for the dermatologist? A study comparing the results of three systems. *Br J Dermatol* 2007;157:926–33.
15. Huang A, Chang W-Y, Hsieh C-H, *et al*. Evaluation of a computer-aided skin cancer diagnosis system for conventional digital photography with manual segmentation. *SPIE Medical Imaging*; International Society for Optics and Photonics, 2014.
16. Hoffmann K, Gambichler T, Rick A, *et al*. Diagnostic and neural analysis of skin cancer (DANAOS). A multicentre study for collection and computer-aided analysis of data from pigmented skin lesions using digital dermoscopy. *Br J Dermatol* 2003;149:801–9.
17. Platt JC. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*; Citeseer, 1999.
18. Deng Y, Manjunath BS. Unsupervised segmentation of color-texture regions in images and video. *IEEE Trans Pattern Anal Mach Intell* 2001;23:800–10.
19. Dietenbeck T, Alessandrini M, Friboulet D, *et al*. CREASEG: a free software for the evaluation of image segmentation algorithms based on level-set. *Image Processing (ICIP), 2010 17th IEEE International Conference on*; IEEE, 2010.
20. Cao Y, Li D, Tavanapong W, *et al*. Parsing and browsing tools for colonoscopy videos. *Proceedings of the 12th Annual ACM International Conference on Multimedia*; ACM, 2004.
21. Wang Y, Zhou Y, Yang J, *et al*. JSEG based color separation of tongue image in traditional Chinese medicine. *Progress in Pattern Recognition, Image Analysis and Applications*; Springer, 2004: 503–8.
22. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996;1:30–46.
23. Jaccard P. The distribution of the flora in the alpine zone. *New Phytologist* 1912;11:37–50.
24. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988:837–45.
25. R: A language and environment for statistical computing [program], 2013.
26. Chen CS, Sierra H, Cordova M, *et al*. Confocal microscopy-guided laser ablation for superficial and early nodular basal cell carcinoma: a promising surgical alternative for superficial skin cancers. *JAMA Dermatol* 2014;150:994–8.
27. Korotkov K, Garcia R. Computerized analysis of pigmented skin lesions: a review. *Artif Intell Med* 2012;56:69–90.
28. Garnavi R, Aldeen M, Celebi ME. Weighted performance index for objective evaluation of border detection methods in dermoscopy images. *Skin Res Technol* 2011;17:35–44.
29. Silveira M, Nascimento JC, Marques JS, *et al*. Comparison of segmentation methods for melanoma diagnosis in dermoscopy images. *IEEE J Selected Top in Signal Process* 2009;3:35–45.