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 Chen13

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, 80.07% and 81.47% for general practitioners, and 88.43% and 85.57% for JSEG. The possibility of direct onsite computation application for physicians other than dermatology specialists when assessing melanocytic and non-melanocytic skin lesions.

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

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 morbidity 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. In the study, a dermatologist manually segmented the images for analysis. Regarding the concerns of subjectivity and consistency of the manually generated borders, many automatic border detection methods have been developed, such as the JSEG algorithm, contrast enhancement and clustering algorithms, with different evaluation metrics. However, many of these algorithms were developed to approximate the ground truth borders, which were also determined by dermatologists subjectively. 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. 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. 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 auto-segmentation 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. 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 macro lens (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. 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.

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
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,
 and tested them by using a diverse data set of 769 images with 19 categories of histological diagnoses.
 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 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, colonoscopy images, tongue images and skin lesion images.

Overlapping and variability

The average or comparable extraction results obtained by multiple dermatologists can be considered more reliable.
 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.

2. Area overlap percentage or Jaccard index, as defined in Eq. (1), between each GP and the gold standard

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

where \(S_{\text{GP}}\) and \(S_{\text{Gold standard}}\) denote the lesion areas determined by the borders drawn by GP and the gold standard, respectively, and \(S_{\text{GP}} \cup S_{\text{Gold standard}}\) 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.

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
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. All statistical analyses were performed using R V.3.1.0.

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.

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

| All lesions                | 347 | 100.00   | 199/147   | 50.97           |

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 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 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+D+C) 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.

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).
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 0.65, 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.

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 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.

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

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

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

GP, general practitioner.
discrepancies between the two groups of physicians. The 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 gener-
ous inclusion criteria, which represents skin lesions
encountered in daily practice, with each lesion given a
definite histopathological diagnosis. SKINCAD per-
formed 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 perfor-
mance with colour different backgrounds.

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

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

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.1
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 examina-
tion, even by expert dermatologists.26 Therefore, sub-
jectivity is always a concern, and automated
segmentation may not always be applicable to these clin-
cal situations. There remained no unified results avail-
able for comparing all the tested segmentation
algorithms due to differences in the ground-truth defini-
tions and evaluation metrics.27–29 When utilising con-
tventional digital photographs for the analysis of melanocytic
and non-melanocytic lesions, the images usually consist
of more colours than the dermoscopic images of melano-
cytic 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 unsuccess-
sful autosegmentation.14 In addition, our software
provides an easy-to-use interface for manual segmenta-
tion. Users are able to complete the process in less than
2 min for each lesion, including the manual segmenta-
tion process and the probability score computation.
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 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
1Graduate Institute of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan
2Department of Dermatology, E-Da Hospital, I-Shou University, Kaohsiung, Taiwan
3Research Center for Adaptive Data Analysis, Center for Biomarkers and Translational Medicine, and Graduate Institute of Biomedical Engineering, National Central University, Jhongli, Taiwan
4Department of Dermatology, Kaohsiung Municipal Ta-Tung Hospital, Kaohsiung, Taiwan
5Department of Dermatology, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan
6Department of Family Medicine, E-Da Hospital, I-Shou University, Kaohsiung, Taiwan
7Graduate Institute of Adult Education, National Kaohsiung Normal University, Kaohsiung, Taiwan
8Chiu Kun-Hsing Dermatology Clinic, Kaohsiung, Taiwan
9Huang Yung-Hsueh Clinic, Changhua, Taiwan
10Health Management Center, Far Eastern Memorial Hospital, New Taipei City, Taiwan
11Department of Family Medicine, Taipei City Hospital, Renai Branch, Taipei, Taiwan
12Department of Psychology, National Chengchi University, Taipei, Taiwan
13Department 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-008-008-009, 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.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.
REFERENCES