Available online at www.sciencedirect.com
ScienceDirect
Computerized Medical Imaging and Graphics 32 (2008) 566–579
Computerized Medical Imaging and Graphics
www.elsevier.com/locate/compmedimag

An improved Internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm

Hitoshi Iyatomi [a,b,*], Hiroshi Oka [c], M.Emre Celebi [d], Masahiro Hashimoto [b], Masafumi Hagiwara [e], Masaru Tanaka [f], Koichi Ogawa [a]

[a] *Department of Electronic Informatics, Hosei University Faculty of Engineering, Japan*
[b] *Department of Dermatology, Keio University School of Medicine, Japan*
[c] *Hiro Clinic, Japan*
[d] *Department of Computer Science, Louisiana State University in Shreveport, United States*
[e] *Department of Information and Computer Science, Keio University Faculty of Science and Technology, Japan*
[f] *Department of Dermatology, Tokyo Women's Medical University Medical Center East, Japan*

Received 19 September 2007; accepted 19 June 2008

Abstract

In this paper, we present an Internet-based melanoma screening system. Our web server is accessible from all over the world and performs the following procedures when a remote user uploads a dermoscopy image: separates the tumor area from the surrounding skin using highly accurate dermatologist-like tumor area extraction algorithm, calculates a total of 428 features for the characterization of the tumor, classifies the tumor as melanoma or nevus using a neural network classifier, and presents the diagnosis. Our system achieves a sensitivity of 85.9% and a specificity of 86.0% on a set of 1258 dermoscopy images using cross-validation.

Keywords: Dermoscopy; Computer-aided diagnosis (CAD); Melanoma; Tumor extraction; Neural network

1. Introduction

In the past few decades, the incidence of malignant melanoma has increased gradually in most parts of the world. In Australia, the incidence of melanoma is now approaching 50 cases per 100,000 population [1]. Although advanced malignant melanoma is often incurable, early-stage melanoma can be cured in many cases, particularly before the metastasis phase. For example, patients with a melanoma less than or equal to 0.75 mm in thickness have a good prognosis and their 5-year survival rate is greater than 93% [2]. Therefore, early detection is crucial for the reduction of melanoma-related deaths.

It is often difficult to distinguish between early-stage melanoma and Clark nevi with the naked eye, especially when small lesions are involved. Dermoscopy or epiluminescence light microscopy (ELM) was introduced to improve the accuracy in the diagnosis of pigmented skin lesions (PSLs) [3]. Stolz et al. developed a convenient diagnosis scheme for dermoscopy images called the ABCD rule [4]. This method quantifies the characteristics of PSLs such as asymmetry, border sharpness, color variegation and the presence/absence of various differential structures and makes a diagnosis based on the total score. A systematic review covering Medline entries from 1983 to 1997 revealed that dermoscopy had 10–27% higher sensitivity [5]. However, dermoscopic diagnosis is often subjective and is therefore associated with poor reproducibility and low accuracy especially in the hands of inexperienced dermatologists. The use of dermoscopy, the accuracy of expert dermatologists in diagnosing melanoma is estimated to be about 75–84% [1,6].

Several groups have developed automated analysis procedures to overcome these problems and reported high levels of diagnostic accuracy [7–17]. Table 1 shows an overview of these studies. Rubegni et al. [11] achieved a sensitivity (SE: melanoma detection accuracy) of 94.3 % and a specificity (SP: benign

* Corresponding author at: Department of Electronic Informatics, Hosei University Faculty of Engineering, 3-7-2 Kajino-cho, Koganei, 184-8584 Tokyo, Japan. Tel.: +81 42 387 6217; fax: +81 42 387 6381.
E-mail address: iyatomi@hosei.ac.jp (H. Iyatomi).

0895-6111/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compmedimag.2008.06.005

Table 1
Comparison of classification performance for malignant melanomas

Source	Author	Year	Segmentation method	Classifier	Total # images	Mel. [a](%)	Dys. [b](%)	SE (%)	SP (%)	Comment
[9]	Ganster et al.	2001	Thresholding + color clustering	kNN	5363	2	19	73	89	
[10]	Elbaum et al.	2001	Thresholding	Linear	246	26	45	100	85	
[11]	Rubegni et al.	2002	Thresholding	ANN	550	36	64	94.3	93.8	
[12]	Hoffman et al.	2003	Clustering + region growing	ANN	2218	22	7	–	–	AUC = 0.844
[13]	Blum et al.	2004	–	Logistic	837	10	11	82.3	86.9	
[18]	Oka et al.	2004	Thresholding	Linear	247	24	76	87.0	93.1	Internet-based
[14]	Burroni et al.	2005	Thresholding	Linear	174	22*	78	71.1	72.1	*Only *in situ*
[15]	Seidenari et al.	2005	–	Linear	459	21	17	87.5	85.7	AUC = 0.933
[16]	Menzies et al.	2005	Semi − auto + manual	Logistic	2420	16	25	91	65	
[17]	Celebi et al.	2007	Region growing	SVM	564	16	55	93.3	92.3	

[a] Percentage of melanomas in the dataset.
[b] Percentage of dysplastic nevi in the dataset.

detection accuracy) of 93.8% on 350 cases of nevi and 200 cases of melanoma using an artificial neural network (ANN). Blum et al. [13] reported a SE of 83.3% and a SP of 86.9% on 753 cases of nevi and 84 cases of melanomas with a logistic regression model. Recently, Celebi et al. [17] achieved a SE of 93.3% and a SP of 92.3% on 476 cases of nevi and 88 cases of melanomas with an improved border detection method and a support vector machine classifier with radial basis function (RBF) kernel. However, several problems have persisted with these software-based approaches. For example, results of these studies are not comparable because of the different image sets used in each one. In addition, these studies were designed to develop a screening system for new patients using standalone systems and therefore they have not been opened to the public.

In 2004, we developed a prototype Internet-based melanoma screening system [18]. The URL of the site has changed and it is now http://dermoscopy.k.hosei.ac.jp. When one uploads a dermoscopy image and the associated clinical data, the system extracts the tumor area, calculates the tumor characteristics and reports a diagnosis based on linear discriminant analysis. The system then registers the uploaded image, the associated clinical data and the diagnosis result into a database. Our preliminary Internet-based system achieved a SE of 87.0% and a SP of 93.1% on 188 Clark nevi and 59 melanomas. Since we made this system open to the public, we have identified several issues that would make the system more practical. We have thus focused on the following topics: (1) expansion of the image database for building a classifier, (2) development of a more accurate tumor area extraction algorithm, (3) extraction of more discriminative diagnostic features, (4) development of an effective classification model, and (5) reduction of the system response time.

Collecting many dermoscopy images for building a classifier is the most important issue ensuring system accuracy and generality. However, this is not a trivial task because in order to obtain the diagnosis information, dermatologists usually need histopathological tests or long-term clinical follow-up. To address this issue, our system is designed to store uploaded dermoscopy images into our database. When we have collected a large number of dermoscopy images with the associated diagnoses, we plan to use them in the development of a more accurate classification model.

Diagnostic accuracy depends greatly on the accurate extraction of the tumor area. Since the late 1990s, numerous solutions that address this issue have been reported [19–30]. A notable

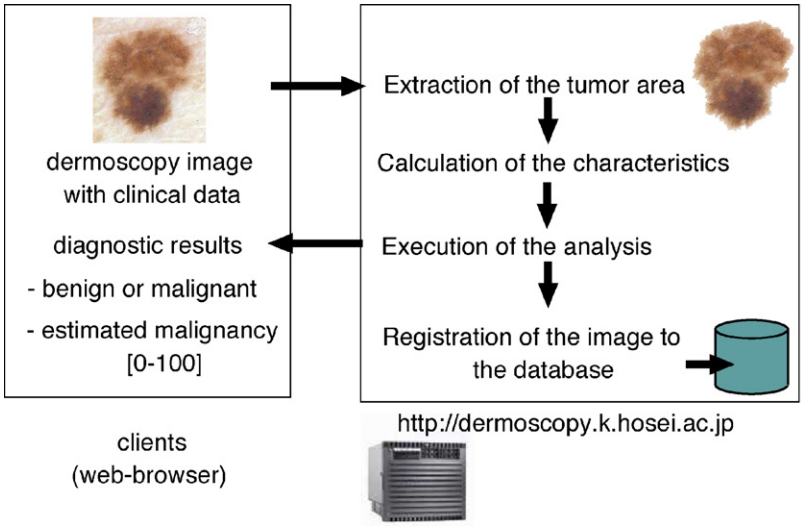


Fig. 1. Overview of the Internet-based screening system.

problem with these studies is that the computer-extracted regions were often smaller than the dermatologist-drawn ones, resulting in the area immediately surrounding the tumor, an important feature in the diagnosis of melanoma, being excluded from the subsequent analysis [1,4,25]. Therefore, there is need for developing a more accurate tumor area extraction algorithm that produces results similar to those determined by the dermatologists.

With regard to the choice of the classifier, despite the fact that simple comparison of classification performance cannot be made, nonlinear classification models such as ANNs [7,11,12], logistic regression models [13,16], and support vector machines [17] have been used.

In this study, we describe the development of a more practical Internet-based melanoma screening system with a dermatologists-like tumor area extraction algorithm [27]. We used a total of 1258 dermoscopy images from three countries and extracted a total of 428 image features from each image. We carefully selected effective features for diagnosis and built an ANN classifier to realize a more practical web-based diagnostic system (Fig. 1).

In addition, a globally accessible Internet-based system requires a well-designed architecture and appropriate user interface facilities. Thus, we revised the internal processing scheme and interfaces to reduce the processing time. Note that our system is a diagnosis support system for dermatologists inexperienced with dermoscopy or physicians of different specialties. The results of our system, therefore, should be considered supplementary.

2. Materials

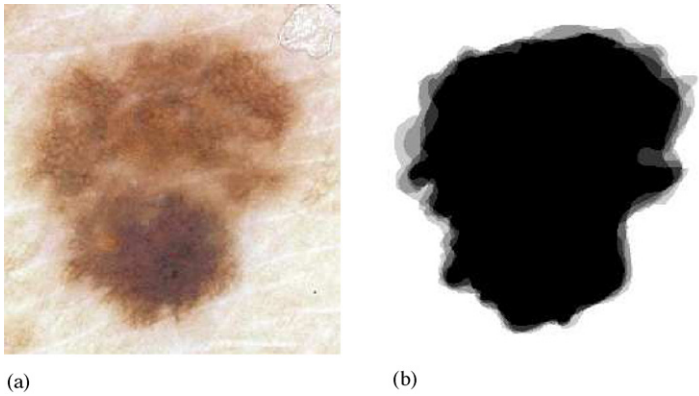
Device calibration to compensate for various imaging conditions, such as different types of dermoscopy, magnification factors, lighting conditions, etc. is crucial in the development of a reliable system. An expert dermatologists would perform the same diagnosis on a particular case even in different imaging conditions. Since scale and color calibration is not feasible in a web-based system, we followed an alternative approach. We developed an accurate classifier using effective and invariant image features extracted from a large and diverse image set.

Digital dermoscopy images of PSLs were collected from four university hospitals (University of Naples, Italy; University of Graz, Austria; Vienna University, Austria; Keio University, Tokyo). These were 24-bit JPEG images with a typical resolution of 768 × 512 pixels. Since we had no control over the image acquisition and camera calibration, we extracted scale-invariant features from images and dermoscopy images that satisfied at least one of the following criteria were omitted from the study: (i) the tumor does not fit entirely within the image frame, (ii) the tumor is part of an acral or mucosal area, and (iii) presence of too much hair. This selectivity was necessary in order to ensure accurate border detection and reliable feature extraction.

We used three different datasets that fulfilled the above criteria:

- *Dataset-A*: 247 dermoscopy images—188 Clark nevi and 59 melanomas (including 23 melanoma *in situ*) from University of Naples and Graz.
- *Dataset-B*: Dataset-A plus 56 Reed nevi and 16 melanomas from University of Naples and Graz (319 dermoscopy images: 188 Clark nevi, 56 Reed nevi and 75 melanomas, including 23 melanoma *in situ*).
- *Dataset-C*: Dataset-B plus 816 melanocytic nevi and 123 melanomas from Keio University and Vienna University (1258 dermoscopy images: 1060 melanocytic nevi and 198 melanomas).

All of the cases in Dataset-A, Dataset-B and some of the cases in Dataset-C were diagnosed based on histopathological examination of biopsy material. The remaining cases in Dataset-C were diagnosed clinically by several expert dermatologists or long-term clinical follow-up. Dataset-A is the dataset from which our prototype system [18] was built and Dataset-B is described in Ref. [27]. Dataset-C was used for developing an accurate classifier. Sensitivity and specificity were used as the evaluation criteria for diagnostic accuracy. We also plotted the receiver operating characteristic (ROC) curve to examine the classifier performance under varying conditions. The diagnostic performance was also quantified by the area under the ROC curve (AUC) measure.



(a) (b)

Fig. 2. Example of tumor areas manually determined by five dermatologists: (a) Dermoscopy image (Clark nevus) and (b) areas manually determined by five dermatologists.

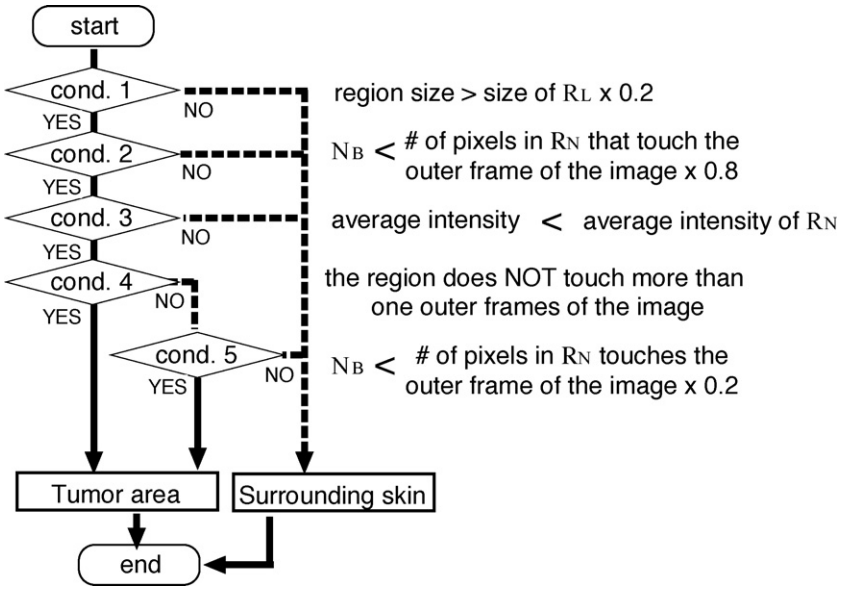


Fig. 3. Flowchart of the tumor area selection phase. R_L: The largest region; R_N: The region that touches the outer frame of the dermoscopy image with the longest boundary; N_B: Number of pixels in the region, which touches the outer frame of the dermoscopy image.

3. Tumor area extraction from surrounding skin

3.1. Definition of the tumor area

For a quantitative evaluation of a tumor area extraction method, we first need the definition of a ground-truth tumor area. Guillod et al. [23,24] used 25 dermoscopy images and found that tumor areas manually extracted by dermatologists were not consistent. Therefore, in order to achieve statistically significant results, we prepared a large number of manual extraction results – 319 images (Dataset-B) × 5 expert dermatologists – and evaluated them quantitatively. Five dermatologists, with an average of 11 years of experience manually determined the borders of all tumors using a tablet computer. In our previous research [27], we compared the extraction results from the 5/5 medical doctor (5/5 MD) area (the region that is selected by all five dermatologists) to the 1/5 MD area (the region that is selected by at least one dermatologist) and evaluated the standard deviation (S.D.) of the selected area as well as the *precision* and *recall*. In that study, we concluded that the area extracted by two or more dermatologists (2/5 MD area) could be taken as the standard tumor area (STA).

Fig. 2 shows a sample of a dermoscopy image and the corresponding manually extracted areas. The black area represents the area selected by all five dermatologists and the gray one is that selected by at least one dermatologist. We evaluated the extraction results using the precision and recall measures:

$$\text{precision}\,(\%) = \frac{\text{correctly extracted area in pixels}}{\text{extracted area in pixels}} \times 100. \quad (1)$$

$$\text{recall}\,(\%) = \frac{\text{correctly extracted area in pixels}}{\text{tumor area in pixels}} \times 100. \quad (2)$$

Note that "tumor area" stands for the standard tumor area and "correctly extracted area" is the overlap between the STA and the extracted area.

3.2. Dermatologist-like tumor area extraction

In our preliminary study we observed that conventional algorithms could mostly extract the tumor areas. However, the extracted areas were often smaller than those determined by dermatologists. Therefore we used our "dermatologist-like" tumor area extraction algorithm [27] that combines both pixel-based and region-based methods and introduces a region-growing approach that aims to bring the automatic extraction results closer to those determined by dermatologists. We evaluated the tumor extraction results by comparing them to the STA.

The "dermatologist-like" tumor area extraction algorithm consists of four phases: (1) initial tumor area decision, (2) regionalization, (3) tumor area selection, and (4) region growing.

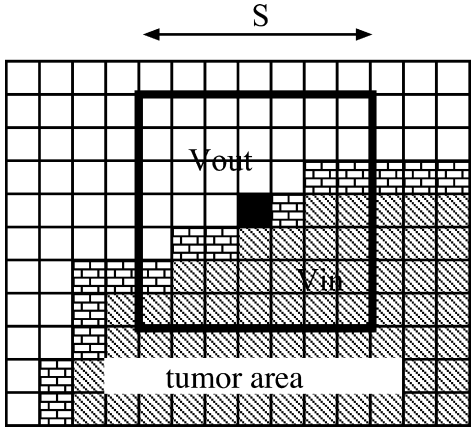


Fig. 4. Notation used in the region-growing algorithm.

(1) *Initial tumor area decision phase*: In this phase a tentative tumor area was selected with a statistical pixel-based thresholding method [31]. Here, we briefly summarize this method. This method uses two filtering operations before the selection of a threshold. First, the image was processed with a Gaussian filter to eliminate the sensor noise. Then, the Laplacian filter was applied to the image and the pixels in the top 20% of the Laplacian values were selected. Only these selected pixels were used to calculate a threshold. The threshold was decided by maximizing the inter-group variance according to a user defined property, such as intensity, red, green or blue [32]. Ganster et al. [9] reported that the blue channel of the RGB color model achieved the best thresholding results and we have also confirmed this in our previous research. Therefore, in this phase, the thresholding was performed on the blue channel and the darker area was taken as a tentative tumor area.

(2) *Regionalization phase*: Because many small isolated regions were created in the previous phase, these needed to be merged in order to obtain a single or sometimes a few small tumor areas. First, a unique region number was assigned to each connected region. Second, a region smaller than ζ% of the image size was combined with the adjacent larger region that shares the longest boundary. This phase makes it possible to manipulate the image as an assembly of regions.

(3) *Tumor area selection phase*: Tumor areas were experimentally determined by selecting appropriate areas from the segmented regions. After consulting with dermatologists, we developed a rule-bases algorithm (illustrated in Fig. 3) based on over 2000 dermoscopy images. The main objective of this phase is to eliminate undesired surrounding shadow areas sometimes produced by the narrow imaging field of the dermoscopy.

The regions that fulfilled the conditions shown in Fig. 3 were selected as the tumor area. In the figure, R_{L}, R_{N}, and N_{B} indicate the largest region, the region that touches the outer frame of the image with the longest boundary, and the number of pixels in the region which touches the outer frame of the image, respectively.

(4) *Region-growing phase*: Based on quantitative experiments that use borders manually drawn by several dermatologists, we have observed that the areas extracted by dermatologists were generally larger than those determined by computer-based methods. Hence in this phase, the extracted tumor area was expanded along the pre-defined border by a region-growing algorithm in order to bring it closer to the area selected by dermatologists.

Fig. 4 illustrates the notation used in the region-growing algorithm. This method traverses the border of the initial tumor using a window of $S \times S$ pixels. When the color properties of the inner V_{in} and outer V_{out} regions are similar, all of the neighborhood pixels are considered as part of the tumor area. In concrete terms, suppose that we have a pixel on the border of the tumor (i.e. the black pixel in Fig. 4) and an $S \times S$ area as part of the tumor area. When condition (3) is satisfied, all pixels in the window are considered as part of the tumor area. This procedure is performed on each and every border pixel (indicated by the brick pattern in Fig. 4). This modification makes the tumor size larger and the border of the tumor is redefined:

$$V_{\mathrm{in}} \times \xi_{\min} \leq V_{\mathrm{out}} < V_{\mathrm{in}} \times \xi_{\max}. \tag{3}$$

Note that $\xi_{\min}$ and $\xi_{\max}$ are the region-growing parameters that need to be determined experimentally. This procedure is repeated iteratively until the size of tumor becomes stable. In other words, the procedure is repeated until there remains no border pixels that satisfies (3).

3.3. Performance comparison with different tumor area extraction algorithms

In a previous study, we evaluated our dermatologist-like tumor area extraction algorithm from a clinical perspective using manually determined borders from five expert dermatologists. Because we have not evaluated this algorithm from an engineering point of view, we compare it with some novel techniques and various state-of-the-art methods. First, we compare our algorithm with novel methods using the STA (gold standard defined by dermatologists):

- Conventional thresholding method [31] (Method A).
- Thresholding + K-means clustering in the $YC_{\mathrm{b}}C_{\mathrm{r}}$ color space (Method B).
- Thresholding + K-means clustering in the HSV color space (Method C).
- Average of borders manually determined by 10 non-medical individuals.
- Our "dermatologist-like" tumor area extraction method.

As a result of preliminary experiments, the following parameter values were used: $S = 7$, $\zeta = 1.00$, $\xi_{\min} = 1.02$, and $\xi_{\max} = 1.07$. These values were experimentally determined using the images in Dataset-B. Method A utilizes a statistical thresholding operation [32] with Gaussian and Laplacian filtering as a pre-processing step. The distance function L of the K-means algorithms used in Methods B and C are defined as $L_{\mathrm{B}} = \sqrt{\delta_Y^2 + \delta_{C_{\mathrm{r}}}^2 + \delta_{C_{\mathrm{b}}}^2}$ and $L_{\mathrm{C}} = \sqrt{\delta_H^2 + \delta_S^2 + \delta_V^2}$, respectively. Here, δ_X refers to the Euclidean distance between feature (X) of target pixel and average feature ($\bar{X}$) of the intermediate cluster. Note here, $\delta_H = \min(H - \bar{H}, H - \bar{H} + 2\pi)$. Based on our preliminary results, introducing a positional term in L deteriorates the results, so we did not use it.

Second, we compare our algorithm with several state-of-the-art methods. Celebi et al. [30] compared the performance of seven recent tumor area extraction algorithms on a set of 90 dermoscopy images using borders manually determined by two dermatologists as the ground-truth. Celebi et al. used the percentage border error [20] as the evaluation criterion:

$$\text{Border error}\,(\%) = \frac{\text{Area}(\text{AutomaticBorder} \oplus \text{ManualBorder})}{\text{Area}(\text{ManualBorder})} \times 100. \tag{4}$$

Table 2
Parameters of neural network classifier in cross-validation test

Items	Values	Items	Values
# of input neurons	5–72	# of training iterations	10–10,000
# of hidden neurons	1–15	Learning coefficient ϵ	0.001–0.2
# of output neuron	1	Momentum coefficient	0–0.9

Here, *AutomaticBorder* is the binary image obtained by filling the computer detected border, *ManualBorder* is obtained by one of the two dermatologists, $\oplus$ is the exclusive-OR operation, and Area(I) denotes the number of pixels in the binary image I.

Celebi et al.'s comparison included optimized histogram thresholding [21], orientation-sensitive fuzzy c-means (OSFCM) [22], gradient vector flow snakes (GVF snakes) [26], our method (dermatologist-like) [27], meanshift clustering [28], modified JSEG [29], and their statistical region merging (SRM) [30] algorithms.

4. Feature extraction, selection and classification

4.1. Feature extraction

After extracting the tumor area, we rotated the tumor object to align its major axis with the Cartesian x-axis. Then, we calculated a total of 428 image-related objective features with reference to the ABCD rule [4]. This rule refers to asymmetry, border sharpness, color variegation, and differential structures of the lesion. The D of ABCD dermoscopy rule evaluates the existence of dermoscopic features such as atypical pigment networks, branched streaks, structureless areas, dots, and globules. It is desirable to calculate features to represent these structures directly. However, extracting these features is often difficult because of the vast variety of dermoscopy images and the highly subjective definitions of these criteria. Argenziano et al. reported that the inter-observer agreement for these features was not high even among expert dermatologists [6]. Therefore, in this study, we extracted texture features such as energy, moment, entropy and correlation of intensity channel over the entire tumor area to quantify the D component.

The calculated features can be roughly categorized into color (140), symmetry (80), border (32) and texture (176) properties. Numbers in parentheses indicate those of calculated image features.

As color-related features, a total of 140 parameters were calculated: minimum (min), average (ave), maximum (max), standard deviation (S.D.) and skewness (skew) values in the RGB and HSV color spaces, respectively (subtotal 30) for the whole tumor area (tumor), periphery of the tumor area (peripheral), difference between the tumor area and the surrounding normal skin (tumor-skin) and difference between peripheral and normal skin (peripheral-skin). In addition, a total of 20 color-related features were calculated; the number of colors in the tumor area and peripheral tumor area in the RGB and HSV color spaces quantized to 8^3 and 16^3 colors, respectively (area 2 × color space 2 × quantized level 2: subtotal 8), the average color of surrounding skin (R, G, B, H, S, V: subtotal 6), and average color differences between the peripheral tumor area and inside of the tumor area (R, G, B, H, S, V: subtotal 6). Note that the peripheral part of the tumor is defined as the region inside the border that has an area equal to 30% of the tumor area and determined by a recursive dilation process applied to the outer border, working inward from the border of the extracted tumor. The ratio of 30% was decided in our preliminary experiments with visual assessment by several dermatologists.

In the symmetry category, a total of 80 features were calculated. We designed 10 intensity (V) threshold values from 5 to 230 with a stepsize of 25. In the extracted tumor area, thresholding was performed and the areas whose intensity was lower than the threshold were determined. From each such area, we calculated eight features: area ratio to original tumor size, circularity, differences of the center of gravity between original tumor (Δ_x, Δ_y), standard deviation of the distribution (σ_x, σ_y) and skewness of the distribution (σ_x^3, σ_y^3).

In order to quantify the border structure, a total of 32 features were calculated. The tumor areas were divided into eight equi-angle regions. In each region, we defined a window of size $S_B \times S_B$ centered on the border of the tumor. In each window, a ratio of color intensity between the inside and outside of the tumor and a gradient of color intensity were calculated in the blue and luminance channels (ratio: B_B and B_L, gradient: B_{Δ_B} and B_{Δ_L}), respectively. These were averaged over the eight equi-angle regions. We calculated four features for eight different window size S_B: 1/5, 1/10, 1/15, 1/20, 1/25, 1/30, 1/35 and 1/40 of the length of the major axis of the tumor object. Note that we used proportionate rated features rather than pre-defined absolute ones in order to handle the images in a scale-invariant manner.

As for the texture features, a total of 176 parameters were calculated. We prepared 11 co-occurrence matrices with distance value δ ranging from 1/2 to 1/64 of the length of the major axis of the tumor object. Based on each co-occurrence matrix, energy, moment, entropy and correlation were calculated in four directions (0°, 45°, 90° and 135°).

These 428 image features were transformed into [0, 1] range using z-score normalization. This feature data was then orthogonalized using the principal component analysis (PCA). In this study, we used two feature sets: the orthogonalized feature set and the original feature set.

4.2. Feature selection

Feature selection is one of the most important steps for developing a classifier. The features used in each classifier were selected by an incremental stepwise method with a hypothesis test of Wilks' lambda [37] using a linear model. This method searches appropriate input parameters one after the other according to the statistical rule. In each step, a statistical F-test was performed and the feature with the highest partial correla-

tion coefficient under $p < 0.05$ (selected feature is statistically effective for the regression) was selected while inefficient (statistically ignorable $p > 0.10$) feature was rejected. This process was continued until a maximum correlation coefficient between outputs of built linear model and the response variables of the model (teach signal) was obtained. Note that this selection method was used for both the orthogonalized feature set and the original dataset. We used these selected features also as input elements of ANN models.

4.3. Diagnosis—ANN and linear classifiers

We used a back-propagation artificial neural network to classify the images based on the calculated features. Although ANNs have excellent learning and function approximation abilities, it is desirable to restrict the number of hidden neurons and input nodes to a minimum in order to obtain a general classification model that performs well on future data [33].

Several training algorithms and refinements for ANNs have been proposed in the literature to enhance the convergence speed and reduce the generalization error of the network. In this study, we used several training algorithms, simple back-propagation algorithm, back-propagation with variable learning rate (VLR), resilient (RES) back-propagation [34], scaled conjugate gradient (SCG) back-propagation [35], and Levenberg–Marquardt (LM) back-propagation [36]. We also tested alternative kernel functions (sigmoid and tangent) and different values for the momentum term.

In our network design, we had only one output node, because our aim was to classify the input as malignant or benign. All nevi, such as Clark nevi, Reed nevi, blue nevi, and dermal nevi, are considered benign. Note that we assigned a training signal of 0.9 and 0.1 to melanoma and benign classes, respectively. If the output of the ANN exceeded the diagnostic threshold θ, we judged the input tumor as malignant.

After the candidates for the input features were determined, we built several ANN classifiers and evaluated them using a leave-one-out cross-validation procedure while changing the number of hidden neurons, training algorithms, training rate ϵ, and number of training iterations. Table 2 shows the training parameters of the ANN classifier used in the cross-validation procedure. When we calculated the performance of the classifier, we moved the diagnostic threshold θ from 0.0 to 1.0 with a stepsize of 0.001 and calculated the SE and SP for each value. For each ANN classifier, we decided the best threshold to be the point where the product of the SE and SP was maximum.

On a separate note, our system provides the screening results not only in the form of "benign" or "malignant", but also as a malignancy score between 0 and 100 based on the output of the ANN classifier. We assigned a malignancy score of 50 to the case where the output of the ANN was θ. For other values, we adjust the score of 0, 20, 80 and 100 according to the output of the ANN of 0, 0.2, 0.8 and 1.0, respectively using linear interpolation. This conversion is based on the assumption that the larger the score of the classifier, the greater the malignancy. Although this assignment procedure is arbitrary, we believe the malignancy score can be useful in understanding the severity of the case.

We also built a linear classifier using the same method as a baseline for the classification performance comparison.

5. Experiments and results

5.1. Tumor area extraction

(1) *Comparison with novel methods*: Figs. 5 and 6 show samples of tumor extraction results. The evaluation results are summarized in Table 3. Method A is the same as the first phase of the proposed method and achieves a very high precision (99%). However, it extracted areas smaller than the STA and therefore the recall remained at low levels. The region-based thresholding methods (Methods B and C) combine Method A and the *K*-means algorithm. The *K*-means and FCM (fuzzy c-means) algorithms, also used in [23,24], need to determine appropriate initial conditions such as the number and shape of the initial clusters. Because they require a large number of clusters in order to achieve certain fitness, they have high computational requirements. Time-consuming procedures are not acceptable, especially for Internet-based systems. Additionally, these methods require other tasks such as selecting proper tumor areas from many regions. In order to overcome these problems, we made Methods B and C to use the result of Method A as the initial clusters for their *K*-means algorithm to reduce the computational time. Segmentation was performed in the YC_bC_r or HSV color spaces (Method B or C) and then the tumor area was determined in the same way as the proposed method. Because two relatively good initial clusters (initial tumor area and surrounding skin area) were given, the processing time was significantly reduced. However, the results of these region-based methods were similar and the methods could not extract important peripheral parts of the tumor as in Method A.

The proposed dermatologist-like algorithm was superior in the extraction of the tumors (precision $= 94.1 \pm 4.5\%$, recall $= 95.3 \pm 5.2\%$) when compared to conventional algorithms and manual extraction results determined by 10 non-medical individuals. The average processing time on Dataset-B (319 images) was approximately 1.5 s on a Pentium 4 2.4 GHz processor.

(2) *Comparison with state-of-the-art methods*: Celebi et al. recently compared seven tumor area extraction algorithms.

Table 3
Comparison of the extraction results for tumor area

Methods	Precision (%)	Recall (%)
(A) Thresholding [31]	99.2	82.8
(B) Thresholding + *K*-means (YC_bC_r)	98.6	84.6
(C) Thresholding + *K*-means (HSV)	99.5	78.8
(D) Average of 10 non-medical individuals	92.1	90.9
(E) Our dermatologist-like	94.1	95.3

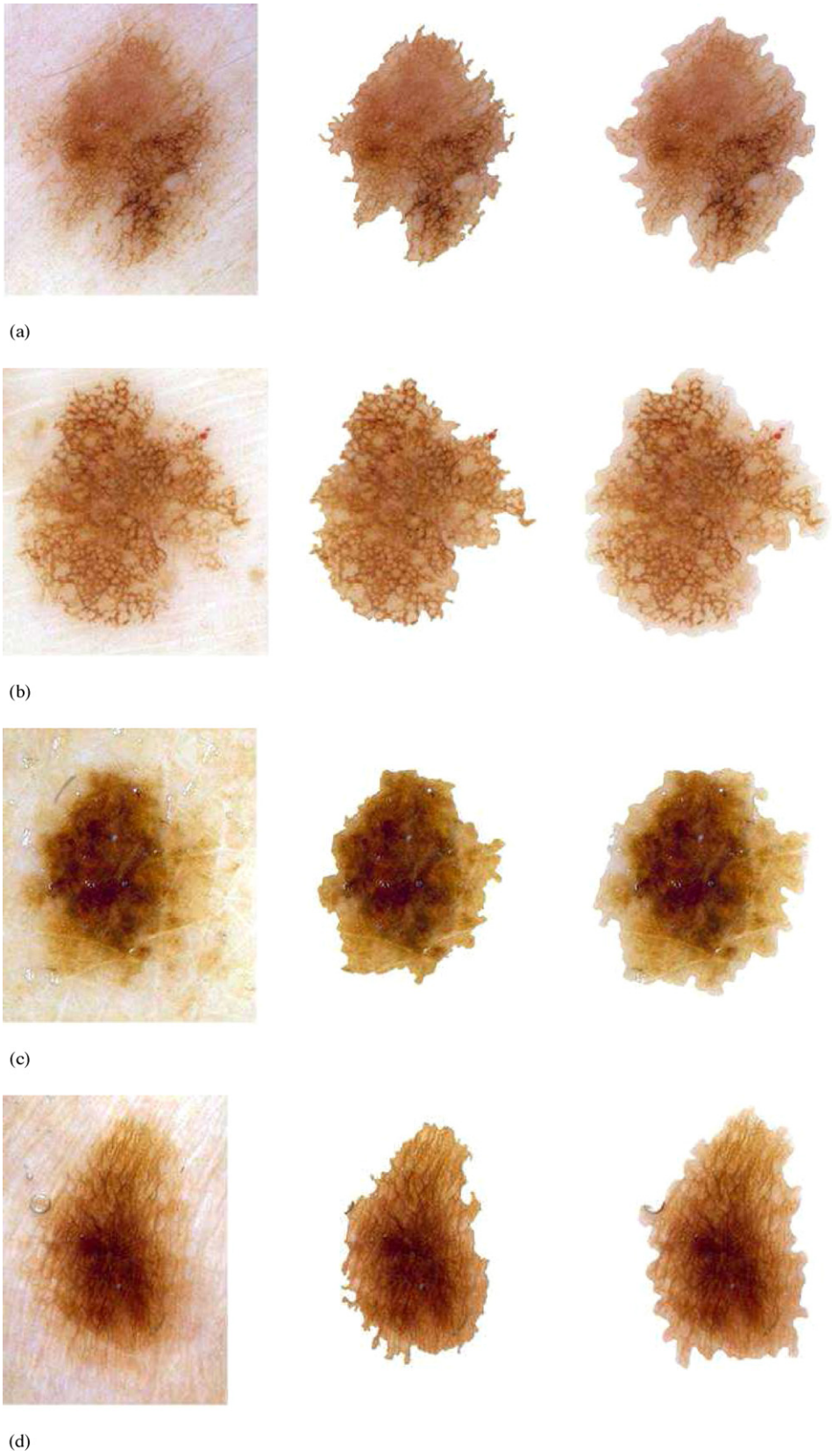


Fig. 5. Examples of tumor area extraction—Clark nevi. (a) Conventional: precision = 100, recall = 74.3; proposed: precision = 98.9, recall = 87.5. (b) Conventional: precision = 100, recall = 79.8; proposed: precision = 98.3, recall = 92.9. (c) Conventional: precision = 100, recall = 76.6; proposed: precision = 95.0, recall = 90.6. (d) Conventional: precision = 99.9, recall = 78.8; proposed: precision = 94.5, recall = 93.6.

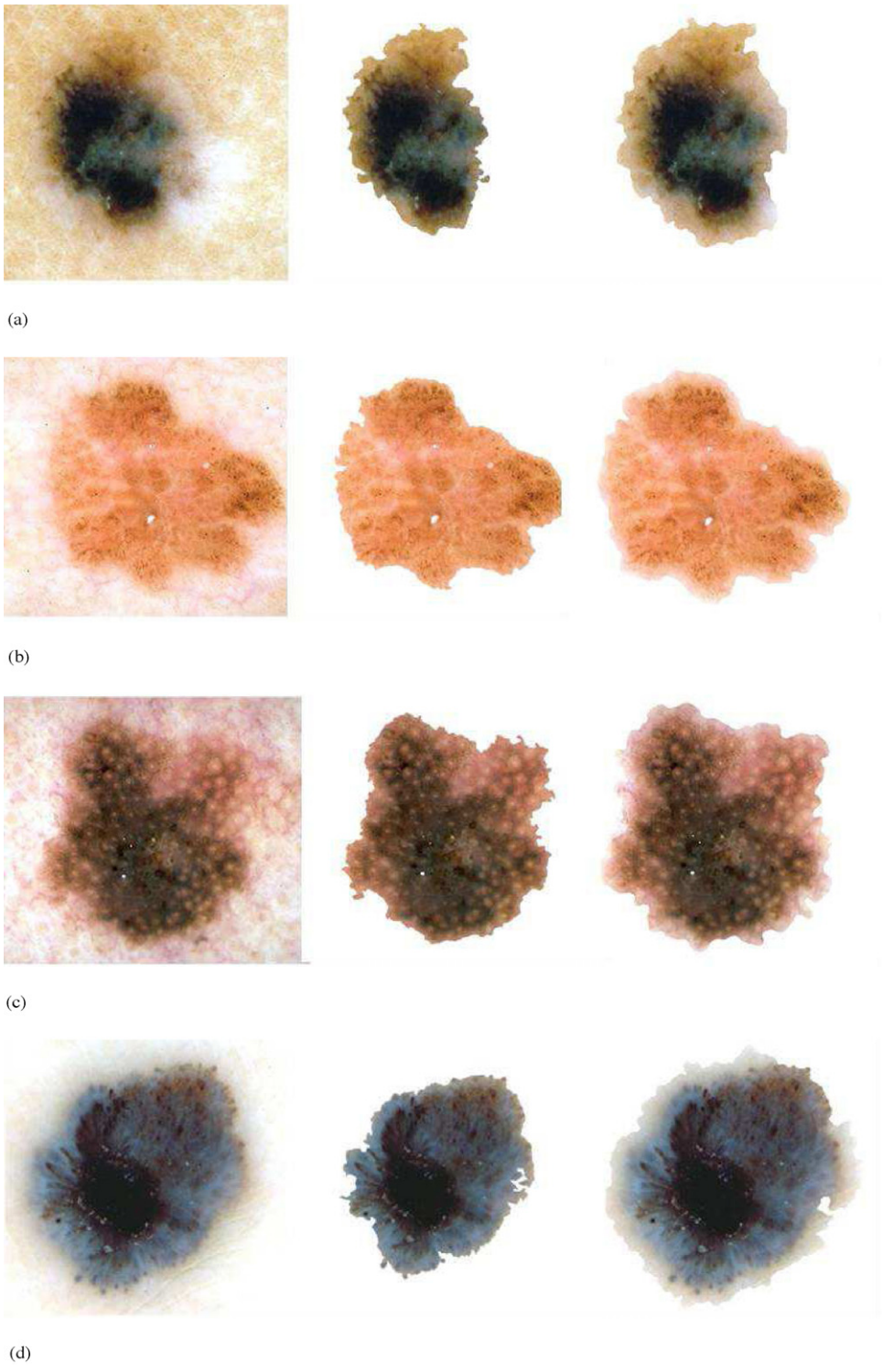


Fig. 6. Examples of tumor area extraction—melanomas. (a) Conventional: precision = 96.2, recall = 76.1; proposed: precision = 84.0, recall = 91.9. (b) Conventional: precision = 100, recall = 88.1; proposed: precision = 97.1, recall = 97.7. (c) Conventional: precision = 100, recall = 75.1; proposed: precision = 97.4, recall = 88.3. (d) Conventional: precision = 100, recall = 84.6; proposed: precision = 98.4, recall = 95.0.

Their results are summarized in Table 4. In their evaluation, our dermatologist-like tumor area extraction algorithm achieved the lowest error in the benign category (mean ± S.D. = 10.66 ± 5.13%) and the second lowest in the overall image set (11.44 ± 6.40%).

5.2. Diagnostic accuracy for evaluation dataset

(1) *Classification performance based on the orthogonalized feature set*: We compared the diagnostic performance of the linear and ANN classifiers on Dataset-C (1258

Table 4
Comparison of the extraction results with recent methods

Method	Year	Benign		Melanoma		Overall	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
Second dermatologist	N/A	8.45	3.75	7.81	3.49	8.28	3.76
Histogram thresholding [21]	1998	12.55	7.27	21.33	15.37	14.79	10.61
OSFCM [22]	1999	23.00	12.61	28.31	15.37	14.79	10.61
GVF snakes [26]	2005	13.44	5.35	19.34	9.34	14.94	7.03
Dermatologist-like[27]	**2006**	**10.66**	**5.13**	**13.70**	**8.93**	**11.44**	**6.40**
Meanshift [28]	2006	11.53	9.74	13.29	7.42	11.98	9.19
Modified JSEG [29]	2007	10.83	6.36	13.74	7.59	11.58	6.77
SRM [30]	2007	10.92	5.73	11.08	6.07	10.96	5.78

Percentage border error statistics (from Ref. [30]). Bold row represents tumor area extraction algorithm in the proposed web-based system (Ref. [27]).

images) (Fig. 7) . The incremental stepwise method selected 72 orthogonalized features from 428 principal components and all selected features were statistically significant ($p < 0.05$). Our linear classifier with 72 input parameters achieved 85.3% in SE, 83.3% in SP and an AUC value of 0.914. Table 5 compares the performance of several ANN and linear models. In this experiment, the basic back-propagation algorithm with constant training coefficients achieved the best classification performance among the tested training algorithms. The ANN classifier with 72 inputs and 6 hidden neurons achieved the best performance of 85.9% in SE, 86.0% in SP, and an AUC value of 0.928.

Table 5
Performance protectcomparison among several ANN and linear models

#in	#hid	hid-func	out-func	ϵ	#train	Algorithm	SE [a]	SP [a]	AUC
12	2	log [b]	Linear	0.02	100	Basic	82.8	68.9	0.835
30	2	log	Linear	0.02	100	Basic	73.2	81.7	0.848
72	2	log	Linear	0.02	100	Basic	70.2	83.5	0.808
30	2	log	Linear	0.02	100	LM	57.1	82.3	0.754
40	2	log	Linear	0.02	100	LM	75.3	74.7	0.807
50	2	log	Linear	0.02	100	LM	63.6	84.5	0.775
72	2	log	Linear	0.02	100	LM	61.6	88.0	0.748
20	2	tan [c]	Linear	0.02	100	Basic	77.3	77.0	0.839
30	2	tan	Linear	0.02	100	Basic	76.8	77.4	0.843
72	2	tan	Linear	0.02	100	Basic	66.7	86.5	0.790
20	2	tan	Linear	0.02	100	LM	71.7	81.7	0.837
72	2	tan	Linear	0.02	100	LM	65.2	89.3	0.789
72	2	tan	log	0.02	100	Basic	68.7	79.9	0.766
72	2	tan	tan	0.02	100	Basic	67.7	79.9	0.734
72	2	log	log	0.02 [d]	100	VLR	80.3	72.6	0.833
72	3	log	log	0.02 [d]	100	VLR	81.8	69.4	0.823
72	6	log	log	0.02 [d]	100	VLR	77.8	71.6	0.809
72	2	log	log	0.02	100	RES	78.8	81.6	0.855
72	6	log	log	0.02	100	RES	75.8	77.2	0.830
72	2	log	log	0.02	100	SCG	74.2	77.2	0.820
72	6	log	log	0.02	100	SCG	75.3	71.2	0.793
72	2	log	log	0.02	100	LM	69.7	79.2	0.802
72	6	log	log	0.02	100	LM	58.6	80.9	0.738
72	2	log	log	0.02	100	Basic	81.8	85.6	0.893
72	5	log	log	0.02	100	Basic	88.4	81.7	0.915
72	6	log	log	0.02	100	Basic	88.9	81.1	0.914
72	8	log	log	0.02	100	Basic	88.4	82.8	0.926
72	10	log	log	0.02	100	Basic	87.9	82.9	0.923
7 2	**6**	**log**	**log**	**0.02**	**30**	**Basic**	**85.9**	**86.0**	**0.928**
72			N/A			**Linear**	**85.3**	**83.3**	**0.914**

The bold values are referred to typical results discussed in the Sections 5.2 and 6.2.
[a] The value at which SE × SP reaches maximum.
[b] sigmoid [0, 1] function.
[c] arctangent [−1, 1] function.
[d] Not constant. Decreasing linearly to 1/100 of initial value.

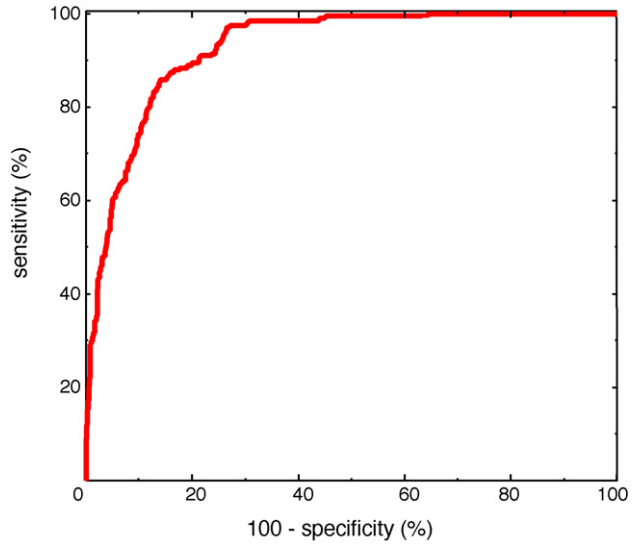


Fig. 7. ROC curve for Dataset-C (1258 images).

Note that the kernel of the hidden neurons and the output neuron were of sigmoid type. The training coefficient and the number of iterations were 0.02 and 30, respectively. Introducing the momentum term boosted the convergence rate at the expense of reduced diagnostic accuracy.

(2) *Classification performance on the original feature set*: In order to show the relationship between input (image features) and output (diagnostic result), we built a linear classifier using the original feature set, without performing PCA. The incremental stepwise method selected 19 features from 428 image features with 27 steps. Some parameters were excluded during the input selection process, because these parameters were highly correlated with some others. The selected 19 features were statistically significant ($p < 0.05$).

These features and the diagnostic accuracy of linear classifier are summarized in Table 6. A linear classifier with 5 image features (step 9) achieved 76.3% in SE, 75.7% in SP, and an AUC value of 0.836 using a leave-one-out cross-validation strategy. Also, a linear classifier with 19 features (step 27) achieved 72.7% in SE, 84.8% in SP, and an AUC value of 0.862.

5.3. Improvement of availability

Our prototype Internet-based diagnosis system [18] was composed of server side Java programs to exploit this language's high degree of compatibility with the Internet-server and the database management system. However, the computational time required by the tumor area extraction and feature calculation modules was

Table 6
Selected image features by the stepwise input selection strategy and the corresponding classification results

Step	addID [a]	exID [b]	Category	Features	#input	SE [c]	SP [c]	AUC
1	374		Texture	Moment, 90°, $\delta = 1/5.7$ of image size	1	60.6	80.6	0.729
2	71		Color	Min, red, peripheral	2	67.7	76.7	0.775
3	59		Color	Average, saturation, normal skin	3	72.7	74.0	0.806
4	34		Color	S.D., red, tumor	4	69.7	76.1	0.813
5	43		Color	Average, green, tumor–normal skin	5	73.7	76.1	0.831
6	219		Asymmetry	σ_x^3, $V \leq 30$	6	76.3	73.3	0.833
7	109		Color	S.D., red, peripheral–normal skin	7	84.8	67.2	0.837
8		59			6	75.3	76.2	0.837
9		71			**5**	**76.3**	**75.7**	**0.836**
10	82		Color	Average, blue, peripheral	6	83.8	70.6	0.842
11	213		Asymmetry	Area ratio, $V \leq 30$	7	83.3	70.8	0.844
12	274		Texture	Moment, 45°, $\delta = 1/45$ of image size	8	82.3	70.1	0.840
13	105		Color	Average, red, peripheral–normal skin	9	73.7	79.0	0.843
14	18		Color	Max, hue, tumor	10	73.7	79.3	0.844
15	370		Texture	Moment, 45°, $\delta = 1/5.7$ of image size	11	70.7	82.7	0.845
16		374			10	71.2	82.7	0.845
17	310		Texture	Moment, 90°, $\delta = 1/23$ of image size	11	81.3	72.4	0.845
18	170		Asymmetry	σ_y, $V \leq 180$	12	79.8	74.7	0.847
19	32		Color	# of color, RGB 16^3, peripheral	13	79.3	75.0	0.852
20	100		Color	Skew, intensity, peripheral	14	78.3	77.0	0.853
21		34			13	82.3	70.1	0.840
22	412		Texture	Correlation, 135°, $\delta = 1/2.8$ of image size	14	76.8	79.4	0.859
23	51		Color	S.D., blue, tumor–normal skin	15	76.3	79.8	0.860
24	160		Asymmetry	Δ_y, $V \leq 205$	16	73.7	83.5	0.860
25	408		Texture	Correlation, 90°, $\delta = 1/2.8$ of image size	17	72.7	83.6	0.861
26	206		Asymmetry	Circularity, $V \leq 55$	18	73.7	83.7	0.862
27	354		Texture	Moment, 45°, $\delta = 1/8$ of image size	**19**	**72.7**	**84.8**	**0.862**

The bold values are referred to typical results discussed in the Section 6.2.

[a] ID of added feature.

[b] ID of excluded feature.

[c] The value at which SE × SP reaches maximum.

high. In fact, in the prototype system, the processing of an image could take up to 20 s. Rubegni et al. pointed out this issue and also indicated that our system had no consideration of security [38].

To overcome these problems, we substituted the time-consuming Java code with a faster C code and provided support for SSL (secure sockets layer) encrypted communication [39]. Despite the introduction of the SSL, the processing time for one image was reduced by almost 75% as a result of these modifications. Furthermore, we modified the user interface so as to enable it to perform the time consuming calculations while the user enters the clinical information, thereby almost completely eliminating the actual waiting time. Recently, out-of-context images were submitted to the server and therefore, we request that the users register their name and e-mail address.

6. Discussion

6.1. Dermatologist-like tumor area extraction

Our tumor area extraction algorithm was designed to achieve scale-invariant results. The parameter ζ used in the regionalization phase did not make a major impact in the final extraction results under different scaled images because this parameter defined proportional value to the image size and it was determined empirically as $\zeta = 1.00$. On the other hand, window size ($S = 7$) was numerically defined and decided empirically. Since the dilation process using this parameter in the region-growing phase was performed recursively, this parameter worked well for images with various sizes. Due to image set size limitations, we did not divide these images into separate training and test sets for the determination of the remaining two parameters, i.e. $\xi_{\min}$ and $\xi_{\max}$. Initially, we used a cross-validation strategy to determine the optimal values for these parameters. However, there were no significant differences among different value combinations. Because the extraction performance of this method was high enough and its S.D. was not large when compared to that of the dermatologists, we judged that the adopted parameter values ($\xi_{\min} = 1.02$ and $\xi_{\max} = 1.07$) were appropriate.

The conventional algorithms achieved almost perfect precision but did not provide sufficient recall. This means that these methods could not extract the tumor area sufficiently. In other words, the border of the tumor, which is an important feature in the identification of melanoma, was often not adequately extracted. The proposed algorithm achieved relatively low precision when compared to the conventional methods. This is due to the fact that the borders of some tumors were ambiguous with variability observed in even the dermatologist-selected areas. Given that the S.D. of the tumor areas manually extracted by five dermatologists was 8.9%, the precision of the proposed algorithm can be considered to be high enough and the extracted areas were almost equivalent to those determined by dermatologists. These results support the conclusion that the proposed dermatologist-like method with a region-growing approach achieves results similar to those attained by dermatologists. In addition, our algorithm provided better performance than that by non-medical individuals. Therefore, we feel that a user interface for manual tumor area extraction will not be necessary when we widen the target audience of the system to individuals not trained in medicine.

According to the comparative study conducted by Celebi et al. (Table 4), our tumor area extraction method achieved the best results in the benign category and the second best in the overall image set. Based on the abovementioned results, we can conclude that our algorithm provides accurate and stable extraction results. Our algorithm is very simple and therefore suitable for an Internet-based system from responsiveness and robustness perspectives.

6.2. Diagnostic performance

The ANN classifier achieves a good classification performance (SE = 85.9%, SP = 86.0%, AUC = 0.928) considering that the diagnostic accuracy of expert dermatologists was 75–84% and that of histological tissue examination on difficult cases was as low as 90% [6]. In our experiments, the basic back-propagation algorithm achieved best classification performance and the linear classifier achieved competitive results. In this study, the feature selection was performed by the incremental stepwise method with Wilks' λ hypothesis test because this method selects or rejects a feature with statistical test in each selection step based on a linear model, effective features for linear model were obtained. It is well known that building a classifier with highly correlated parameters is adversely affected by so called multicollinearity, and, in such a case, the system loses accuracy and generality. This input selection method rejects statistically ignorable features during incremental selection and therefore, these highly correlated features were automatically excluded from the model. When we used the orthogonalized feature set, this method did not exclude any features during the selection. On the other hand, as we can see from Table 6, some features were excluded from the regression model with the original feature set. With this functionality, the developed model is considered to be reliable.

ANNs with different training algorithms did not achieve better performance as expected. Several training parameters such as training iterations and coefficients were selected based on the results of the basic back-propagation. If we find an effective training parameter set or effective input features for each model, we expect that the performance will be better. As shown in Table 6, classification performance of the linear classifier based on the original image feature set was lower (AUC = 0.862) than that obtained by using PCA-based dataset (AUC = 0.914), but relationship between input and output was clear. The linear classifier based on only five features (step = 9: texture, three color, and asymmetry features) achieved reasonable results (SE = 76.3%, SP = 75.7%, AUC = 0.836) and based on these results we confirmed again that color features are important for discriminating melanoma. If we can extract more discriminating features from the image, better classification performance can be obtained. Recent studies on high-level dermoscopic feature extraction include two pilot studies on pigment networks

[40,41] and globules [41] and four systematic studies on dots [42], blotches [43,44], and blue-white areas [45]. We plan to incorporate these kinds of features in the next version of our system.

6.3. The Internet-based screening system and data standardization

We used JPEG images for building the classifier, because they are commonly used in the practice of dermoscopy. The images used in this study are of high quality and it is unlikely that the use of the JPEG format would influence the results significantly.

The total processing time for one image was 3–5 s. In most cases, the system completes processing while the user enters clinical information associated with the image.

Despite the good classification performance obtained, our system has several limitations with regard to the acceptable tumor classes and the condition of the input images. However, these are in line with the limitations imposed in previous studies. At the present, the diagnostic capability of our system does not match that of expert dermatologists, primarily because of the lack of a large and diverse dermoscopy image set. We address this issue by expanding our images set continually with the help of university hospitals and private clinics.

Our Internet-based diagnostic system serves another important purpose as a public dermoscopy image repository. We plan to open our database to public use in the near future with the hope that it will enable objective comparisons to be made among various studies.

Currently, our system is limited for use by dermatologists or physicians only. We would like to widen the target audience with further improvements to the system. Our short-term objective is to design more effective and intuitive features that would enable us to build a more accurate classifier using a large number of dermoscopy images. In addition, we are planning to provide a more detailed output along with the final diagnosis result. If the system were to provide quantitative measures for various clinical criteria, such as asymmetry, border structure, color, differential structures, etc. this would increase its clinical acceptance.

7. Conclusions

In this paper, we have developed an Internet-based melanoma screening and data collection system. Key components of this system are a dermatologist-like tumor area extraction algorithm and an artificial neural network classifier. The proposed dermatologist-like tumor area extraction algorithm was superior in extraction performance (precision = 94.1%, recall = 95.3%) when compared to conventional methods, results of manual extraction by 10 non-medical individuals, and various state-of-the-art methods described in the literature. Our neural network classifier with improved tumor area extraction results achieved very competitive results (SE = 85.9%, SP = 86.0%, AUC = 0.928) using a cross-validation test on a set of 1258 dermoscopy images. With the Internet connection, anyone who has a dermoscopy image can use our screening system from all over the world. In its current form, our system should not be considered as an alternative for dermatologists, but it should be viewed as an effective diagnosis support system that has the capability of finding early-stage melanomas.

Acknowledgments

This research was partially supported by the Ministry of Education, Science, Sports and Culture Grant-in-Aid for Young Scientists program (B), 17790788, 2005-2006.

References

[1] Stolz W, Falco OB, Bliek P, Kandthaler M, Burgdorf WHC, Cognetta AB. Color atlas of dermatoscopy, 2nd enlarged and completely revised edition. Berlin: Blackwell publishing; 2002.
[2] Meyskens Jr FL, Berdeaux DH, Parks B, Tong T, Loescher L, Moon TE. Natural history and prognostic factors influencing survival in patients with stage I disease. Cancer 1998;62(6):1207–14.
[3] Soyer HP, Smolle J, Kerl H, Stettnre H. Early diagnosis of malignant melanoma by surface microscopy. Lancet 1987;2:803.
[4] Stolz W, Riemann A, Cognetta AB, Pillet L, Abmayr W, Holzel D, et al. ABCD rule of dermatoscopy: a new practical method for early recognition of malignant melanoma. European Journal of Dermatology 1994;4(7):521–7.
[5] Mayer J. Systematic review of the diagnostic accuracy of dermoscopy in detecting malignant melanoma. The Medical Journal of Australia 1997;167(4):206–10.
[6] Argenziano G, Soyer HP, Chimenti S, Talamini R, Corona R, Sara F, et al. Dermoscopy of pigmented skin lesions: results of a consensus meeting via the Internet. Journal of American Academy of Dermatology 2003;48(5):679–93.
[7] Ercal F, Chawla A, Stoecker WV, Lee HC, Moss RH. Neural network diagnosis of malignant melanoma from color images. IEEE Transactions on Biomedical Engineering 1994;41(9):837–45.
[8] Seidenari S, Pellacani G, Pepe P. Digital videomicroscopy improves diagnostic accuracy for melanoma. Journal of American Academy of Dermatology 1998;39(2):175–81.
[9] Ganster H, Pinz A, Rohrer R, Wilding E, Binder M, Kitter H. Automated melanoma recognition. IEEE Transactions on Medical Imaging 2001;20(3):233–9.
[10] Elbaum M, Kopf AW, Rabinovitz HS, Langley RG, Kamino H, Mihm Jr MC, et al. Automatic differentiation of melanoma from melanocytic nevi with multispectral digital dermoscopy: a feasibility study. Journal of American Academy of Dermatology 2001;44:207–18.
[11] Rubegni P, Cevenini G, Burroni M, Perotti R, Dell'Eva G, Sbano P, et al. Automated diagnosis of pigmented skin lesions. International Journal of Cancer 2002;101:576–80.
[12] Hoffmann K, Gambichler T, Rick A, Kreutz M, Anschuetz M, Grunendick T, 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. British Journal of Dermatology 2003;149:801–9.
[13] Blum A, Luedtke H, Ellwanger U, Schwabe R, Rassner G, Garbe C. Digital image analysis for diagnosis of cutaneous melanoma. Development of a highly effective computer algorithm based on analysis of 837 melanocytic lesions. British Journal of Dermatology 2004;151:1029–38.
[14] Burroni M, Sbano P, Cevenini G, Risulo M, Dell'Eva G, Barbini P, et al. Dysplastic naevus vs. in situ melanoma: digital dermoscopy analysis. British Journal of Dermatology 2005;152:679–84.
[15] Seidenari S, Pellacani G, Grana C. Pigment distribution in melanocytic lesion images: a digital parameter to be employed for computer-aided diagnosis. Skin Research and Technology 2005;11:236–41.
[16] Menzies SW, Bischof L, Talbot H, Gutener A, Avramidis M, Wong L, et al. The performance of SolarScan—an automated dermoscopy image analysis instrument for the diagnosis of primary melanoma. Archives of Dermatology 2005;141(11):1388–96.

[17] Celebi ME, Kingravi HA, Uddin B, Iyatomi H, Aslandogan YA, Stoecker WV, et al. A methodological approach to the classification of dermoscopy images. Computerized Medical Imaging and Graphics 2007;31(6):362–73.
[18] Oka H, Hashimoto M, Iyatomi H, Tanaka M. Internet-based program for automatic discrimination of dermoscopic images between melanoma and Clark nevi. British Journal of Dermatology 2004;150(5):1041.
[19] Green A, Martin N, Pfitzner J, O'Rourke M, Knight N. Computer image analysis in the diagnosis of melanoma. Journal of American Academy of Dermatology 1994;31(6):958–64.
[20] Hance GA, Umbaugh SE, Moss RH, Stoecker WV. Unsupervised color image segmentation with application to skin tumor borders. IEEE Engineering in Medicine and Biology Magazine 1996;15(1):104–11.
[21] Pagadala P. Tumor border detection in epiluminescence microscopy images. MS thesis. Rolla: Department of Electrical and Computer Engineering, University of Missouri; 1998.
[22] Schmid P. Segmentation of digitized dermatoscopic mages by two-dimensional color clustering. IEEE Transactions on Medical Imaging 1999;18(2):164–71.
[23] Guillod J, Schmid-Saugeon P, Guggisberg D, Jean PC, Braun R, Krischer J, et al. Validation of segmentation techniques for digital dermoscopy. Skin Research and Technology 2002;8(4):240–9.
[24] Schmid-Saugeon P, Guillod J, Thiran JP. Towards a computer-aided diagnosis system for pigmented skin lesions. Computerized Medical Imaging and Graphics 2003;27:65–78.
[25] Grana C, Pellacani G, Cucchiara R, Seidenari S. A new algorithm for border description of polarized light surface microscopic images of pigmented skin lesions. IEEE Transactions on Medical Imaging 2003;22(8):959–64.
[26] Erkol B, Moss RH, Stanley RJ, Stoecker WV, Hvatum E. Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes. Skin Research and Technology 2005;11(1):17–26.
[27] Iyatomi H, Oka H, Saito M, Miyake A, Kimoto M, Yamagami J, et al. Quantitative assessment of tumour area extraction from dermoscopy images and evaluation of the computer-based methods for automatic melanoma diagnostic system. Melanoma Research 2006;16(2):183–90.
[28] Melli R, Grana C, Cucchiara R. Comparison of color clustering algorithms for segmentation of dermatological images. Proceedings of the SPIE Medical Imaging 2006;6144:3S1–19S319.
[29] Celebi ME, Aslandogan YA, Stoecker WV, Iyatomi H, Oka H, Chen X. Unsupervised border detection in dermoscopy images. Skin Research and Technology 2007;13(4):454–62.
[30] Celebi ME, Kingravi HA, Iyatomi H, Aslandogan YA, Stoecker WV, Moss RH, et al. Border Detection in Dermoscopy Images Using Statistical Region Merging. Skin Research and Technology 2008;14(3):347–353.
[31] Tanaka T, Murase Y, Ueya T. Classification of Sarcomas based on texture features. Journal of Medical Imaging Technology 2001;19(1):33–41.
[32] Otsu N. An automatic threshold selection method based on discriminant and least square criteria. Transactions of IEICE 1988;63:349–56.
[33] Reed R. Pruning algorithms—a survey. IEEE Transactions on Neural Networks 1993;4(5):740–7.
[34] Riedmiller M. Advanced supervised learning in multilayer perceptrons—from backpropagation to adaptive learning algorithms. Computer standards and Interfaces 1994;16(3):265–78.
[35] Hagan MT, Demuth HB, Beale MH. Neural network design. Boston: PWS Publishing; 1996.
[36] Hagan MT, Menhaj M. Training feedforward networks with the Marquardt algorithm. IEEE Transactions on Neural Networks 1994;5(6):989–93.
[37] Everitt BS, Dunn G. Applied multivariate data analysis. London: Edward Arnold; 1991. p. 219–20.
[38] Rubegni P, Burroni M, Sbano P, Andreassi L. Digital dermoscopy analysis and internet-based program for discrimination of pigmented skin lesion dermoscopic images. British Journal of Dermatology 2005;152(2):395–6.
[39] Oka H, Iyatomi H, Hashimoto M, Tanaka M. Reply to "Digital dermoscopy analysis and internet-based program for discrimination of pigmented skin lesion dermoscopic images". British Journal of Dermatology 2006;154(3):570–1.
[40] Fleming MG, Steger C, Zhang J, Gao J, Cognetta AB, Pollak I, et al. Techniques for a structural analysis of dermatoscopic imagery. Computerized Medical Imaging and Graphics 1998;22(5):375–89.
[41] Caputo B, Panichelli V, Gigante GE. Toward a quantitative analysis of skin lesion images. Studies in Health Technology and Informatics 2002;90:509–13.
[42] Yoshino S, Tanaka T, Tanaka M, Oka H. Application of morphology for detection of dots in tumor. Proceedings of SICE Annual Conference 2004;1:591–4.
[43] Stoecker WV, Gupta K, Stanley RJ, Moss RH, Shrestha B. Detection of asymmetric blotches (asymmetric structureless areas) in dermoscopy images of malignant melanoma using relative color. Skin Research and Technology 2005;11(3):179–84.
[44] Pellacani G, Grana C, Cucchiara R, Seidenari S. Automated extraction and description of dark areas in surface microscopy melanocytic lesion images. Dermatology 2004;208(1):21–6.
[45] Celebi ME, Kingravi HA, Aslandogan YA, Stoecker WV. Detection of blue-white veil areas in dermoscopy images using machine learning techniques. Proceedings of the SPIE Medical Imaging 2006: 1861–8.

Hitoshi Iyatomi is a research associate of Hosei University. He was born in Tokyo, Japan in 1976. He received his BE and ME degrees in Electrical Engineering in 1998 and 2000 and PhD degree in Open and Environmental Systems in 2004 from Keio University, respectively. During 2000–2004 he was employed by Hewlett Packard Japan. He was COE (center of excellence) researcher at Keio University from 2003–2004. His research interests include image understanding, machine learning, and medical image analysis.

Hiroshi Oka is a director of Hiro Clinic, Kawaguchi, Japan. He was born in Okayama, Japan in 1971. He received his medical and PhD degrees in Keio University School of Medicine in 1996 and 2005, respectively. He is working as a dermatologist and a cosmetic surgeon.

M. Emre Celebi received his BSc degree in computer engineering from the Middle East Technical University (Ankara, Turkey) in 2002. He received his MSc and PhD degrees in computer science and engineering from the University of Texas at Arlington (Arlington, TX, USA) in 2003 and 2006, respectively. He is currently an assistant professor in the Department of Computer Science at the Louisiana State University in Shreveport (Shreveport, LA, USA). His research interests include medical image analysis, color image processing, content-based image retrieval, and open-source software development.

Masahiro Hashimoto is a medical doctor of Keio University. He was born in Wakayama, Japan in 1981. He received his medical degree in Keio University, School of Medicine in 2006.

Masafumi Hagiwara is a professor of Keio University. He was born in Yokohama, Japan, in 1959. He received the BE, ME, and PhD degrees in electrical engineering in Keio University, Yokohama Japan, in 1982, 1984 and 1987, respectively. In 1987, he became a research associate of Keio University. Since 1995, he has been an associate professor. From 1991 to 1993 he was visiting scholar at Stanford University. His research interests include soft computing. Dr. Hagiwara is a member of the IEEE, IEICE, IEE of Japan, IPSJ (Information processing society of Japan) and JNNS.

Masaru Tanaka is professor of dermatology, Tokyo Women's Medical University Medical Center East from 2007. He was born in Tokyo, Japan in 1958. He received his MD and PhD degrees at Keio University in 1984 and 1992, respectively. He studied abroad in Cardiff, UK from 1993 to 1995. He was assistant professor from 1996 and associate professor from 1999 at the Department of Dermatology, Keio University School of Medicine. His research interests include dermatopathology, dermoscopy and image analysis.

Koichi Ogawa graduated from Keio University, College of Engineering, Department of Electrical Engineering in 1980. In 1982, completed masters at Keio, became assistant in College of Medicine. In 1990, became instructor in College of Medicine at Keio. Became associate professor at Hosei's College of Engineering in 1991, full professor in 1998. Work on image processing and radiometry. Doctorate in Engineering. Received award for paper from Japan Medical Imaging Society in 1989, awards from Japan Nuclear Medicine Society in 1991. Member IEEE, SNM, Japan Medical Radiology Society, Japan Nuclear Medicine Society, etc.