An Internet-based Melanoma Screening System - Supported Acral Volar Lesions

Hitoshi Iyatomi, M.Emre Celebi, Hiroshi Oka and Masaru Tanaka

Abstract— **In this paper, we present an Internet-based melanoma screening system that newly supports acral volar lesions. A half of asian melanomas are from these areas and they show completely different appearance from other lesions. Our screening system is accessible from all over the world and diagnoses dermoscopy images within 3-5 sec based on a neural network classifier for non-acral lesions or newly integrated linear classifier for acral volar lesions. Our system achieves a sensitivity of 85.9% and a specificity of 86.0% on a set of 1258 non-acral dermoscopy images and a sensitivity of 93.3% and a specificity of 91.1% on a set of 199 acral volar dermoscopy images using cross-validation.**

I. Introduction

Although advanced malignant melanoma is often incurable, early-stage melanoma can be cured in many cases, particularly before the metastasis phase. 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 was introduced to improve the accuracy in the diagnosis of pigmented skin lesions (PSLs) [1]. However, dermoscopic diagnosis is often subjective and is therefore associated with poor reproducibility. Despite the use of dermoscopy, the accuracy of expert dermatologists in diagnosing melanoma is estimated to be about 75-84% [2].

Several groups have developed automated analysis procedures to overcome these problems and reported high levels of diagnostic accuracy [3]-[12]. Table I shows an overview of these studies. 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 such background we developed a prototype of fully automated Internet-based melanoma screening system at our university[8]. 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. The system then registers the uploaded image, the associated clinical data and the diagnosis result into a database. Current our system has more sophisticated automatic tumor area extraction algorithm, which achieved similar extraction performance with expert dermatologist [13] and a neural network classifier. Our latest back-propagation artificial neural network classifier (ANN) achieved high classification performance of a sensitivity of 85.9% and a specificity of 86.0% on a set of 1258 dermoscopy images (1060 melanocytic nevi and 198 melanomas) using cross-validation [14].

On the other hand, in non-white populations, almost half of the melanomas are found in acral volar areas and nearly 30% of melanomas affect the sole of the foot [15]. Saida et al. also reported that melanocytic nevi are also frequently found in their acral skin and approximately 8% of Japanese have melanocytic nevi on their soles.They reported that about 90% of melanomas in this area have the parallel ridge pattern and 70% of melanocytic nevi have the parallel furrow pattern. In fact, the appearance of these acral volar lesions is largely different from PSLs found in other body areas (Fig.1) and accordingly development of specially designed classifier is required for these lesions.

However, automatic detection of the parallel ridge and parallel furrow patterns is often difficult due to the wide variety of dermoscopy images and there had been no published method on computerized classification of this diagnostic category. Recently we succeeded to quantify these patterns and developed a classification model for these lesions [16].

In this paper, we developed new diagnosis classifier for acral volar lesions and integrated it on our Internet-based screening system. We also developed new user-friendly interface makes the users to input the location of the PSL. According to this information the system automatically selects appropriate classifier.

Our system is a diagnosis support system for dermatologists inexperienced with dermoscopy or physicians of a different specialty. The results of our system therefore should be considered subsidiary.

II. Material

A total of 199 acral volar digital dermoscopy images (169 benign and 30 melanomas) were collected from four Japanese hospitals and two European university hospitals. The details are shown in [16]

Dermoscopy images that satisfied at least one of the following criteria were omitted before from the study: (i)

H.Iyatomi is with Faculty of Engineering, Hosei University, 3-7-2 Kajinocho Koganei, 184-8522, Tokyo, Japan (phone: 81-42-387-6217; fax: 81-42-387-6381; e-mail:iyatomi@hosei.ac.jp). M.E.Celebi is with the Department of Computer Science, Louisiana State University in Shreveport, Shreveport, LA 71115 USA (phone: 1-318-795-4281; fax: 1-318-795-2419; e-mail: ecelebi@lsus.edu). H.Oka is with Department of Cosmetic Suergeon, Shiromoto Clinic. M.Tanaka is with Department of Dermatology, Tokyo Women's Medical University Medical Center East.

TABLE I

COMPARISON OF CLASSIFICATION PERFORMANCE FOR MALIGNANT MELANOMAS.

Source	Author	Year	Segmentation method	Classifier	Total # images	Mel.† (%)	Dys. ‡ (%)	SE (%)	SP (%)	comment
[3]	Ganster et al.	2001	Thresholding+color clustering	kNN	5363	2	19	73	89	
[4]	Elbaum et al.	2001	Thresholding	Linear	246	26	45	100	85	
[5]	Rubegni et al.	2002	Thresholding	ANN	550	36	64	94.3	93.8	
[6]	Hoffman et al.	2003	clustering+region growing	ANN	2218	22	7	-	-	AUC=0.844
[7]	Blum et al.	2004	-	Logistic	837	10	11	82.3	86.9	
[8]	Oka et al.	2004	Thresholding	Linear	247	24	76	87.0	93.1	Internet-based
[9]	Burroni et al.	2005	Thresholding	Linear	174	22⋆	78	71.1	72.1	⋆: only *in situ*
[10]	Seidenari et al.	2005	-	Linear	459	21	17	87.5	85.7	AUC=0.933
[11]	Menzies et al.	2005	Semi-auto+manual	Logistic	2420	16	25	91	65	
[12]	Celebi et al.	2007	Region growing	SVM	564	16	55	93.3	92.3	
[14]	Iyatomi et al.	2008	clustering+region growing "Dermatologist-like"[13]	ANN	1258	16	84*	85.9	86.0	AUC=0.928 Internet-based

†: Percentage of melanomas, ‡: Percentage of dysplastic nevi, ∗: Percentage of all melanocytic nevi in the data set.

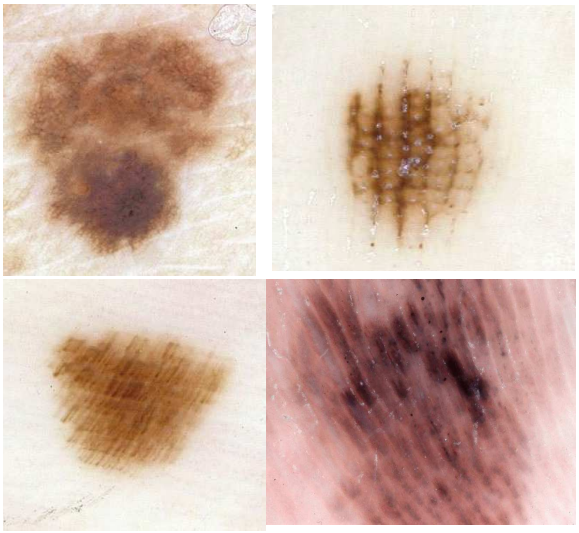

Fig. 1. Examples of dermoscopy images
from top-left: Clark nevus (common benign), acral benign.
from bottom-left acral fibrillar pattern (benign) and acral melanoma

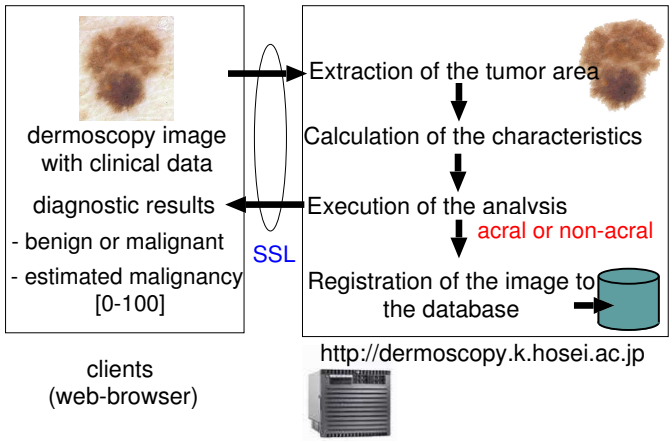

Fig. 2. Overview of the Internet-based screening system.

the tumor does not fit entirely within the image frame, (ii) the tumor is part of mucosal area, and (iii) presence of too much hair. This selectivity was necessary in order to ensure accurate border detection and reliable feature extraction.

All cases were histopathologically or clinically diagnosed materials and these results were used as the gold standard (training data) for building the classifier.

III. BUILDING THE CLASSIFIER

The appearance of acral volar lesion is largely different from PSLs from other body areas. In our previous research, we found that discriminating acral volar melanomas require a specifically designed classifier, which focuses more on asymmetry and textural properties of the image rather than only color. According to this, we newly develop a classifier for these lesions and integrate it to our system.

A. *Tumor area extraction from surrounding skin*

Detection of appropriate tumor area is one of the important step for computer-aided melanoma screening. We used our "dermatologist-like" tumor area extraction algorithm [13] 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 confirmed our algorithm achieves results similar to those attained by dermatologists and better than non-medical individuals [13] and several automated novel techniques [14]. Celebi *et al.* [17] recently compared seven tumor area extraction algorithms. In their evaluation, our algorithm achieved the lowest border error in the benign category (mean±SD = 10.66±5.13%) and the second lowest in the overall image set (11.44±6.40%).

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 this study and an Internet-based system from responsiveness and robustness perspectives.

B. 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 consideration of several diagnostic criteria for dermoscopy. The calculated features can be roughly categorized into color (140), symmetry (80), border (32) and texture (176) properties. Number in parenthesis indicates that of calculated image features. The detailed description of these features are shown in our previous paper [14].

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). Each principal component (PC) is used as a candidate for input element for the classifier.

C. Feature selection

Feature selection is also very important issue for developing a classifier. The features used in the classifier was selected by an incremental stepwise method with a hypothesis test of Wilks' lambda [18] using linear model. This method searches appropriate input parameters one after the other according to the statistical rule. In each step, statistical F-test is performed and the feature with the highest partial correlation coefficient under $p < 0.05$ (selected feature is statistically effective for the regression) is selected while inefficient (statistically ignorable $p > 0.10$) feature is 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. We define the number of the selected PCs as $\#in_{\max}$.

D. Diagnosis and Evaluation

1) Developing the classifier: In our previous study, linear model accurately detect parallel ridge and parallel furrow patterns with abovementioned feature set, so we used it in this study, too. We also developed logistic regression model for a performance comparison.

We assigned a training signal of 1 and 0 to melanoma and benign classes. If the output of the developed model exceeded the diagnostic threshold θ, we judged the input tumor as being malignant.

2) Performance evaluation: We developed several linear models and logistic regression models with different number of input elements ($1 \leq \#in \leq \#in_{\max}$). We evaluated these classifiers using a leave-one-out cross-validation procedure.

Sensitivity (SE) and specificity (SP) were used as the evaluation criteria for diagnostic accuracy. For clinical purposes, it is important to keep the SE at high level. 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.

3) Malignancy score: Our system have provided 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 classifier [14]. We adopt same mechanism at this time. We assigned a malignancy score of 50 to the case where the output of the classifier was θ. For other values, we adjust this score of 0, 20, 80, and 100 according to the output of the classifier of 0, 0.2, 0.8 and 1.0, respectively using linear interpolation.This conversion is based on the assumption that the larger score of the classifier is, the more malignancy is.

IV. Results and Discussion

A. Classification performance

The classification performance of the linear models was summarized in TableII. The incremental stepwise method selected 45 PCs ($\#in_{\max}$=45) and all selected PCs were statistically significant ($p < 0.05$).

The linear classifier with 45PCs achieved 100% in SE, 95.9% in SP, and an AUC value of 0.996. In general, 45 inputs for 199 data is too much. We performed 10-fold cross-validation and it showed almost equivalent results of AUC=0.991. Because we used orthogonalized feature set, the risk of so called "multicolineality" can be ignored. With above reasons, we considered overfitting was not occurred in this test.

On the other hand, the classifier with only 5 PCs, those were selected first in the feature selection and they were important PCs for detection of parallel ridge and parallel furrow patterns [16], achieved very competitive diagnostic accuracy of 93.3% in SE, 91.1% in SP, and an AUC value of 0.933. The detailed components of these PCs are shown in [16]. Note that parameters chosen early in the stepwise feature selection were thought to be more important for the classification because the statistically most significant parameters were selected in each step.

The diagnosis performance of logistic regression models are almost equivalent to the that of linear models. When the number of the input elements is 5,10,30 and 45, the corresponding AUC is 0.931, 0.947, 0.944 and 0.998, respectively.

Since the system with smaller number of the input should have high generality in common and linear model is the most simple architecture, we integrate this 5-input linear classifier on our server. Fig.3 shows the ROC curve for our latest screening system: for (1) non-acral and (2) acral lesions.

B. New interface

We modify user interface for providing easy access. Our new interface requires the user to input the position of the lesion corresponds to the submitted image. With this information, the system automatically judge whether the submitted image is from acral volar or not. On the other hand, the system store this information to own database for further follow-up or other purpose.

Our system requires around 3-5 sec to make final result without network latency. Current system performs the time consuming calculations while the user enters the clinical

TABLE II

CLASSIFICATION PERFORMANCE OF ACRAL VOLAR MELANOMA

(LINEAR MODEL)

$\#in$	SE (%)†	SP (%)†	AUC
1	93.3	65.1	0.833
3	83.3	87.6	0.917
5	93.3	91.1	0.934
10	93.3	92.3	0.956
20	96.7	94.1	0.983
30	96.7	95.3	0.993
45	100	95.9	0.996

†: The SE and SP values shown are those that have the maximum product.

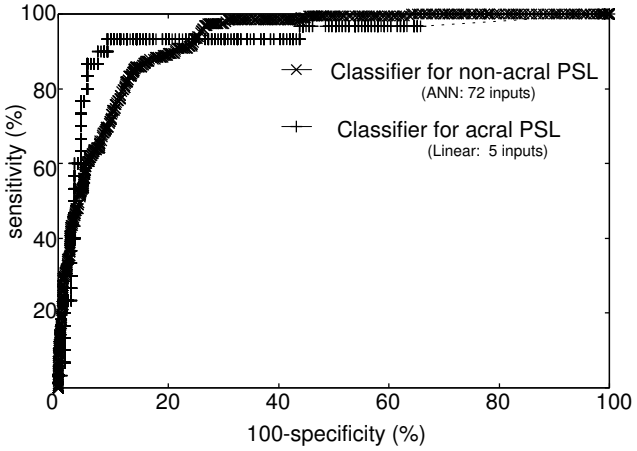

Fig. 3. ROC curves of the classifiers of our screening system

information thereby almost completely eliminating the actual waiting time.

V. CONCLUSIONS

In this paper we developed new diagnostic classifier for acral volar lesion and integrated to our Internet-based melanoma screening and data collection system. Supporting acral volar lesions greatly improve the system availability, because in non-white populations almost half of melanomas are found in these area. With the internet connection, everyone who have dermoscopy can use our screening system from all over the world. In current format, our system cannot be an alternative of dermatologist, but it would be efficient diagnosis support system and have a capability for finding early stage hidden patients in the future.

ACKNOWLEDGEMENTS

This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (C), 20591461, 2008-2010.

REFERENCES

[1] H.P.Soyer, J.Smolle, H.Kerl, H.Stettnre, "Early diagnosis of malignant melanoma by surface microscopy," Lancet, Vol.2, p.803, 1987.

[2] G.Argenziano, H.P.Soyer, S.Chimenti, R.Talamini, R.Corona, F.Sara et al., "Dermoscopy of pigmented skin lesions: Results of a consensus meeting via the Internet," Journal of American Academy of Dermatology, Vol.48, No.5, pp.679 693, 2003.

[3] H.Ganster, A.Pinz, R.Rohrer, E.Wilding, M.Binder, H.Kitter, "Automated melanoma recognition, " IEEE Trans. on Medical Imaging Vol.20, No.3, pp.233-239, 2001.

[4] M.Elbaum, A.W.Kopf, H.S.Rabinovitz, R.G.Langley, H.Kamino, M.C. Mihm Jr. et al., "Automatic differentiation of melanoma from melanocytic nevi with multispectral digital dermoscopy: a feasibility study," Journal of American Academy of Dermatology, Vol.44, pp.207-218, 2001.

[5] P.Rubegni, G.Cevenini, M.Burroni, R.Perotti, G.Dell'Eva, P.Sbano P et al.,"Automated diagnosis of pigmented skin lesions," International Journal of Cancer, Vol.101, pp.576-580, 2002.

[6] K.Hoffmann, T.Gambichler, A.Rick, M.Kreutz, M.Anschuetz, T.Grunendick et al., "Diagnostic and neural analysis of skin cancer (DANAOS). A multicentre stidy for collection and computer-aided analysis of data from pigmented skin lesions using digital dermoscopy," British Journal of Dermatology, Vol.149, pp.801-809, 2003.

[7] A.Blum, H.Luedtke, U.Ellwanger, R.Schwabe, G.Rassner, C.Garbe, "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, Vol.151, pp.1029-1038, 2004.

[8] H.Oka, M.Hashimoto, H.Iyatomi, M.Tanaka, "Internet-based program for automatic discrimination of dermoscopic images between melanoma and Clark nevi," British Journal of Dermatology, Vol.150, No.5, p.1041, 2004.

[9] M.Burroni, P.Sbano, G.Cevenini, M.Risulo, G.Dell'Eva, P.Barbini et al., "Dysplastic naevus vs. in situ melanoma: digital dermoscopy analysis," British Journal of Dermatology, Vol.152, pp.679-684, 2005.

[10] S.Seidenari, G.Pellacani, C.Grana, "Pigment distribution in melanocytic lesion images: a digital parameter to be employed for computer-aided diagnosis," Skin Research and Technology, Vol.11, pp.236-241, 2005.

[11] S.W.Menzies, L.Bischof, H.Talbot, A.Gutener, M.Avramidis, L.Wong et al., "The performance of SolarScan - An. automated dermoscopy image analysis instrument for the diagnosis of primary melanoma," Archives of Dermatology, Vol.141, No.11, pp.1388-1396, 2005.

[12] M.E.Celebi, H.A.Kingravi, B.Uddin, H.Iyatomi, Y.A.Aslandogan, W.V.Stoecker, R.H.Moss, "A methodological approach to the classification of dermoscopy images," Computerized Medical Imaging and Graphics, Vol.31, No.6, pp.362-371, 2007.

[13] H.Iyatomi, H.Oka, M.Saito, A.Miyake, M.Kimoto, J.Yamagami 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, Vol.16, No.2, pp.183-190, 2006.

[14] H.Iyatomi, H.Oka, M.E.Celebi, M.Hashimoto, M.Hagiwara et al., "An Improved Internet-based Melanoma Screening System with Dermatologist-like Tumor Area Extraction Algorithm," Computerized Medical Imaging and Graphics 2008 (In press).

[15] T.Saida, A.Miyazaki, S.Oguchi, Y.Ishihara, Y.Yamazaki, S.Murase et al, "Significance of dermoscopic patterns in detecting malignant melanoma on acral volar skin, " Arch Dermatol,Vol.140, pp.1233-1238, 2004.

[16] H.Iyatomi, H.Oka, M.E.Celebi, K.Ogawa, G.Argenziano, H.P.Soyer et al, "Computer-based classification of dermoscopy images of melanocytic lesions on acral volar skin," Journal of Investigative Dermatology (In press) 2008.

[17] M.E.Celebi, H.A.Kingravi, H.Iyatomi et al., "Border Detection in Dermoscopy Images Using Statistical Region Merging," Skin Research and Technology, 2008 (In press).

[18] B.S.Everitt, G.Dunn, Applied Multivariate Data Analysis. London: Edward Arnold; p.219-220, 1991.