Four-Class Classification of Skin Lesions with Task Decomposition Strategy

Hagar Maher

10th October 2018

1 Abstract

- This paper proposes a new computer-aided method for skin lesion classification applicable to both melanocytic skin lesions (MSLs) and non-melanocytic skin lesions (NoMSLs).
- Several researchers have developed methods to distinguish between melanoma and nevus, which are both categorized as MSL. However, most of these studies did not focus on NoMSLs such as basal cell carcinoma (BCC) the most common skin cancer and seborrheic keratosis (SK) despite their high incidence rates.
- It is preferable to deal with these NoMSLs as well as MSLs especially for the potential users who are not enough capable of diagnosing pigmented skin lesions on their own such as dermatologists in training and physicians with different expertise.
- They developed a new method to distinguish among melanomas, nevi, BCCs, and SKs.
- Their method calculates 828 candidate features grouped into three categories:
    1. Color
    2. Subregion
    3. Texture
- They introduced two types of classification models:
    1. A Layered model that uses a task decomposition strategy
    2. Flat models to serve as performance baselines
- They tested their methods on 964 dermoscopy images:
    1. 105 melanomas
    2. 692 nevi
    3. 69 BCCs
    4. 98 SKs
- The layered model outperformed the flat models, achieving detection rates of 90.48%, 82.51%, 82.61%, and 80.61% for melanomas, nevi, BCCs, and SKs, respectively.
- They also identified specific features effective for the classification task including irregularity of color distribution.
- The results show promise for enhancing the capability of computer-aided skin lesion classification.

2 Introduction

- Basal cell carcinoma (BCC), the most common skin cancer is rarely fatal, but it destroys surrounding tissue if left untreated. Thus, early detection and appropriate treatment are essential.
- Detection of skin cancers is difficult due to the confusing appearance of wide variety of skin lesions.
- Biopsy provides a definitive diagnosis, however it can cause metastasis and therefore is only allowed based on the premise of following surgical operation within a month.
- In addition, these are invasive operations and make unpleasant experiences to the patient.
- To avoid unnecessary biopsy, several researchers investigated non-invasive computer-aided methods to distinguish melanomas from nevi using dermoscopy images.
- These methods usually consist of three steps:
    1. Border detection of skin tumor
    2. Feature extraction
    3. Classification
- The border detection process finds the border of the tumor in the dermoscopy image, which is essential for accurate skin lesion classification.
- Several methods have been proposed such as dermatologist-like method, SRM, hybrid thresholding, threshold fusion, and so on.
- The feature extraction process obtains discriminating image features that facilitate classification such as general color statistics, contour shape, and texture information.
- Wavelet coefficients that capture color and shape information have also been investigated.
- The classification process determines the type of skin lesions from the extracted image features.
- General pattern classifiers such as linear discriminant classifier, k-NN, artificial neural networks, and SVMs are often used.
- Based on the aforementioned three steps, researchers have improved the automated classification methods. Although there are several limitations, these studies reported superior classification performance compared to experts.

- They developed an internet-based melanoma screening system (current URL is https://dermoscopy.k.hosei.ac.jp) which they continually update to improve accuracy and reliability.
- The abovementioned conventional studies have several problems:
    1. Only limited types of skin lesions are acceptable for classification
    2. The systems do not explain the reasons for the classification results
    3. The systems were developed and evaluated with only ideal condition images and did not consider the condition of test images
- In this paper, they focus on the first issue, i.e. the limitation of applicable skin lesion types.
- Most of the conventional works handled only melanocytic skin lesions (MSLs) such as melanomas and nevi which originate from melanocytes, whereas non-melanocytic skin lesions, (NoMSLs) indicating all the other pigmented skin lesions except MSLs such as basal cell carcinomas (BCCs) and seborrheic keratoses (SKs) have been relatively neglected. This is because melanoma is the most fatal skin cancer and especially difficult to differentiate from nevus even by expert dermatologists.
- Although classification of NoMSLs is considered to be easier than that of MSLs for expert dermatologists, it is not always easy for inexperienced dermatologists or physicians with different expertise. Therefore, if they open the system also for those potential users, as was the case with the abovementioned internet-based system, it is necessary to handle both MSLs and NoMSLs.
- They developed a general border detection algorithm for MSLs and NoMSLs. Finding the border of NoMSLs was a challenging task because they often have unclear borders. With this sophisticated algorithm, they found that a linear classifier with only two image features ('skewness of bright region on the major axis' and 'difference in blue intensity between the peripheral and the normal skin') discriminated MSLs from NoMSLs with performance of 98.0% SE and 86.6% SP.
- They further developed a system to detect melanomas from other MSLs (nevi) and NoMSLs. Using 548 MSL and 110 NoMSL images, the system achieved 97.6% SE and 87.7% SP (89.5% SP for nevi and 79.1% SP for NoMSLs).
- However, this study focused only on discriminating melanomas from all other lesions, thus clumping BCCs with benign skin tumors.
- In this paper, they propose a method to distinguish among four types of skin lesions: melanoma, nevus, BCC, and SK, using a significantly larger dataset.

3 Dataset

In this study, they used 968 digital dermoscopy images categorized into four types: melanoma, nevus, BCC, and SK.

- Melanoma: 105 images (30 from Keio University Hos-pital and 75 from University of Naples and Graz), a malignant melanocytic tumor (MSL), the most fatal skin cancer.
- Nevus: 692 images (448 from Keio University Hospital and 244 from University of Naples and Graz), a benign melanocytic tumor (MSL), often difficult to differentiate from melanomas.
- Basal cell carcinoma (BCC): 69 images (20 from Keio University Hospital and 49 from Tokyo Women's Med-ical University), a malignant non-melanocytic tumor (NoMSL), the most common skin cancer.
- Seborrheic keratosis (SK): 98 images (42 from Keio University Hospital and 56 from Tokyo Women's Medical University), a benign non-melanocytic tumor (NoMSL), which commonly occurs in the elderly and is sometimes confused with melanomas.

These images have different resolutions ranging from 512*384 to 3641*2732. The diagnosis of the skin lesions was determined by histopathological examination or clinical agreement by several expert dermatologists.

4 Method

4.1 Border Detection

- From each skin lesion image, we extracted the border between the tumor and the surrounding normal skin area. Ac-curate border detection usually results in better classification performance.
- Conventional automated methods of border detection mostly focused on only melanocytic skin lesions (MSLs).
- In their previous study, we developed a general border detection algorithm for both MSLs and NoMSLs.
- The core of the algorithm is color thresholding, removal of artifacts such as microscope border and hair, and inclusion of bright area seen specifically in NoMSLs.
- The algorithm outperformed other state-of-the-art methods (dermatologist-like method, SRM, hybrid thresholding, k-means++, and JSEG for NoM-SLs and showed equivalent or better performance for MSLs.

4.2 Feature Extraction

- After determining the border of the tumor, we segmented the skin lesion image into four regions as illustrated in Fig. 1: normal skin, peripheral, central tumor, and whole tumor.

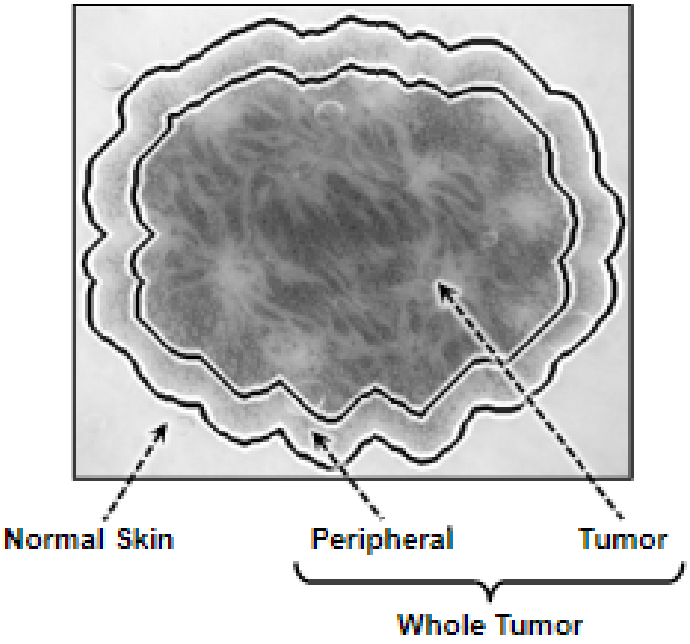

Fig. 1. Four regions in the skin lesion image.

1. The whole tumor consists of all pixels within the extracted border.
2. The normal skin is all pixels on the outside of the border.
3. The peripheral is the first 30% of the whole tumor area, obtained by going inward from the border as in our previous studies
4. The central tumor is obtained by removing the peripheral from the whole tumor.

- For preprocessing, They rotated the images to make the major axis of the whole tumor parallel to its horizontal axis (X-axis).
- They resized the images such that the major axis of the whole tumor was 512 pixels in length due to the disparate image resolutions and to reduce computation time.
- After preprocessing, they calculated 828 candidate image features which are mostly variants of the 428 image features from our previous studies.

- The reason for introducing new features is that the previous 428 features were designed purely for detecting melanomas while in this study we distinguish among four types of skin lesions.
- The 828 features are grouped into the three categories:
    1. Color (300)
    2. Subregion (144)
    3. Texture (384)
- The numbers in the parentheses denote those of the features in the corresponding categories.

4.2.1 Color Related Features

- They calculated 10 statistics (min, max, standard deviation, skewness, entropy
    1. 5%-tile
    2. 25%-tile
    3. 50%-tile
    4. 75%-tile
    5. 95%-tile

    of the intensity of 6 color channels
    1. R: red
    2. G: green
    3. B: blue
    4. H: hue
    5. S: saturation
    6. V: luminance

    for each of the 3 tumor regions
    1. Peripheral
    2. Central tumor
    3. Whole tumor
- This yielded 180 parameters (10 statistics x 6 channels x 3 regions).
- They also calculated the difference in the same 10 statistics on the 6 color channels between central tumor and peripheral and those between peripheral and normal skin area, which yielded 120 parameters (10 statistics x 6 channels x 2 pairs-of-regions).

- They expect these difference-oriented features to be robust over variations of dermoscopy images caused by different photographic conditions.
- In total, there are 300 color related features (180+120).
- The main change made from their previous studies is the adoption of %-tile statistics.
- The reason for using %-tile is that they are expected to be robust over artifacts such as black hairs and shiny bubbles compared to min, mean, or max.

4.2.2 Sub-region Related Features

- Subregion related features describe geometrical distribution of the color.
- First, they divided the central tumor and the peripheral into smaller even subregions as illustrated in Fig. 2.

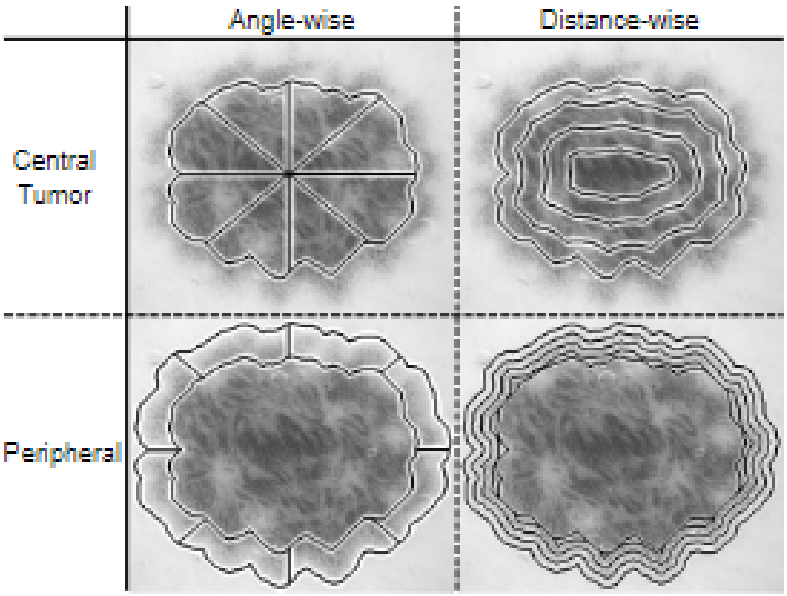

Fig. 2. Subregions of central tumor and peripheral.

- They used 2 types of subdivisions, angle-wise and distance-wise.
    1. The angle-wise is based on the angle from the center of gravity of the central tumor to the edge of the region.

2. The distance-wise is based on the Euclidean distance from the outer border.

- They used 3 numbers of subregions: 4, 8, and 16 for the angle-wise manner and 2, 4, and 8 for the distance-wise manner.
- For each subregion, they calculated 3 statistics (mean, standard deviation, and skewness) on 4 color channels (R, G, B, and S).
- They left out H and V because these two channels did not contribute to classification performance in their reliminary experiments.
- They calculated the standard deviation of these statistics within all subregions. This yielded 144 subregion features (2 target regions x 2 types of subdivisions x 3 numbers of subregions x 4 color channels x 3 statistics for each subregion).
- In their previous studies, they had the asymmetry features to describe geometrical distribution of the color. However, in their preliminary experiments, they determined that the subregion features are more effective for the four-class skin lesion classification.

4.2.3 Texture Related Features

- As for texture related features, they adopted the gray level co-occurrence matrix (GLCM).
- They obtained the GLCMs with the following settings:
  1. 2 target regions (central tumor and whole tumor)
  2. 3 quantization levels (N = 16, 32, and 64)
  3. 4 distances (δ = 1, 2, 4, and 8 pixels)
  4. 4 directions (θ = 0, 45, 90, and 135 from the major axis)
- From each GLCM, they extracted 4 GLCM-statistics
  1. Energy
  2. Correlation
  3. Entropy
  4. Homogeneity
- To make the directional settings (θ) more meaningful, they extracted min, mean, max, and difference (i.e. maxmin) of the abovementioned GLCM-statistics in 4 main directions (θ) as was also recommended in the original literature of the GLCM.
- In total, there are 384 texture features (2 regions x 3 quantization levels x 4 distances x 4 directions (e.g. max) x 4 GLCM-statistics).

- For postprocessing, they normalized all of the 828 features so that they have mean of 0 and variance of 1 over all images in the datasets.
- Note that only a small number of features were selected from the 828 for the classifier development as will be described later.

4.3 Classification

- They introduce the proposed layered model as the primary classification model and the flat model as a performance baseline shown in Figs. 3 and 4, respectively.

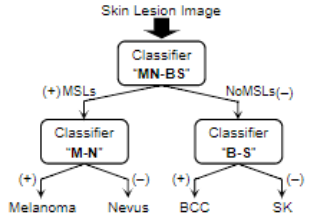

Fig. 3. Overview of the layered model.

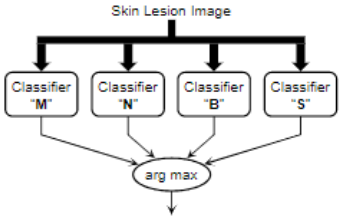

Fig. 4. Overview of the flat model.

- The letters M, N, B, and S in the figures denote melanoma, nevus, BCC, and SK, respectively.

- They used linear classifiers over non-linear ones in order to gain a clear understanding of the relationship between the inputs and the outputs of the models and to facilitate a comparison of classification performance.

  1. Layered model (proposed):
     - The first-step classifier "MN-BS" identifies the input skin lesion as MSL if the output value is greater than the classifier's threshold value or as NoMSL otherwise. These are shown by (+) and () in Fig. 3. If the result is an MSL, the second-step classifier "M-N" distinguishes melanoma from nevus in the same manner by comparing its output value with the threshold value. If the result from the first-step classifier is a NoMSL, the second-step classifier "B-S" distinguishes BCC from SK. The idea of the layered model is to decompose the whole classification task to
       (a) The broad classification of MSL and NoMSL by the "MN-BS".
       (b) The detailed classification of "melanoma and nevus" and that of "BCC and SK" by the "M-N" and "B-S", respectively.

- It may be inferred that the first-step classifier “MN-BS” must have high accuracy because misclassifications at this phase are fatal.
- They designed this model based on the results of theur past studies that distinguishing MSLs from NoMSLs is relatively easy.
- One of the most important steps for classifier development is feature selection. It is well-known that too many features or irrelevant features lead to poor performance and the over-fitting problem. Therefore, it is necessary to select an appropriate subset of features for each of the classifiers “MN-BS”, “M-N”, and “B-S”.
- They adopted Wilks’ Lambda stepwise feature selection method as in their previous studies. This algorithm begins with no selected features and repeats the step of adding or removing a feature one by one iteratively. The feature added is the one which gives the highest increase in linear regression fitness under the F-test (p ¡ 0.05). A feature is removed when it no longer contributes to the linear regression fitness (p ¿ 0.10). This iterative process of adding and removing features continues until no features pass the test for addition or removal.
- After selecting the input features, they trained the linear classifiers. The assigned supervisory outputs were either +1 or 1 as specified by (+) and () in Fig. 3.
- After the training step, they adjusted the threshold values of the three linear classifiers by full search to optimize classification performance, e.g. detection rate of melanoma.
- In their preliminary experiment, they also tested a different layered model which distinguishes cancer (melanoma and BCC) from no-cancer (nevus and SK) at the top level. However, the performance was not satisfactory mainly because of the difficulty in the classification between cancer and no-cancer.

2) Flat models (performance baseline)

- They introduce two types of flat models, namely the “Flat model I” and the “Flat model II” as the performance baseline.
- Each of the flat models has four linear classifiers: “M”, “N”, “B”, and “S” whose output values estimate the presence/absence of the corresponding classes: melanoma, nevus, BCC, and SK, respectively.
- This kind of classification model is typically used for multiclass classification.
- To compare the outputs of the four classifiers, the following score F_i is calculated for each classifier i (e.g. the “M”)

$$F_i = \alpha_i \times (O_i - \xi_i) \qquad (1)$$

Here, O_i is the normalized output value of the classifier i whose standard deviation is 1.
The E_i and α_i are the threshold value and the scaling factor, respectively. The classification result is given by argmaxi F_i. Note that scaling factors used in the flat models are not necessary for the layered model.

- The flat model I and the flat model II are different in how the classifiers possess the features.
- In the flat model I, all classifiers share the same features. they select the features with the Wilks' Lambda stepwise method with the strategy that it improves overall classification performance.
- In the flat model II, each classifier possesses its own features. they select the features specifically effective for each classifier by the Wilks' Lambda stepwise method as well as the layered model.
- For the flat model I, it is necessary to take all four classifiers into consideration when selecting a feature to add or remove in the stepwise method. This issue was addressed in which examined multiclass classification. It suggests two methods: either to optimize 'average' or 'maximum' of the four error reduction amounts associated with the four outputs. They chose the 'average' method because it showed better performance in our preliminary experiments.
- After the feature selection step, they trained the four classifiers for each of the two flat models.
- The supervisory outputs were +1 for the target type (melanoma for the classifier "M") and 1 for the rest (nevus, BCC, and SK for the "M").
- They adjusted the threshold E_i and the scaling factor i shown in equation (1) by means of full search to optimize classification performance.
- The search scope of α_i was empirically defined as $[2^{0.5}; 2^{1.5}]$

5 Results

- Table I shows the top-3 selected features for each classifier of the layered model and the two flat models.

TABLE I
TOP 3 SELECTED FEATURES FOR EACH LINEAR CLASSIFIER

Classifier		First Feature	Second Feature	Third Feature
Layered Model	"MN-BS" (MSL vs NoMSL)	Col: 95%-tile (S:p-n) †	Tex: homogen (ct) [θdif, N64, δ4] ⋆	Col: 25%-tile (H:wt)
	"M-N" (melanoma vs nevus)	Col: σ(V:p)	Sub: $\sigma(\mu$(R:ct)) [angle-16]	Col: 25%-tile (G:p-n)
	"B-S" (BCC vs SK)	Sub: $\sigma(\kappa$(S:ct)) [angle-4] ⋄	Tex: entropy (ct) [θdif, N16, δ8]	Sub: $\sigma(\sigma$(R:ct)) [distance-4]
Flat Model I	shared by all classifiers	Tex: homogen (wt) [θdif, N64, δ4]	Col: 75%-tile (S:p-n)	Col: σ(V:p)
Flat Model II	"M" (melanoma vs else)	Col: σ(V:p)	Sub: $\sigma(\mu$(R:ct)) [angle-16]	Col: κ(G:p-n)
	"N" (nevus vs else)	Tex: homogen (wt) [θdif, N64, δ4]	Col: 5%-tile (S:t-p)	Sub: $\sigma(\mu$(R:ct)) [angle-16]
	"B" (BCC vs else)	Col: 25%-tile (H:wt)	Sub: $\sigma(\sigma$(S:ct)) [angle-16]	Col: 5%-tile (B:p-n)
	"S" (SK vs else)	Col: 95%-tile (S:p-n)	Sub: $\sigma(\mu$(B:ct)) [distance-8]	Tex: entropy (wt) [θmin, N64, δ8]

Feature categories: Col: color, Sub: subregion, Tex: texture.
Target regions: ct: central tumor, wt: whole tumor, p: peripheral, n: normal skin area, ct-p: difference between central tumor and peripheral.
Statistics: μ: mean, σ: standard deviation, κ: skewness.
†: Difference in 95%-tile of saturation (S) between peripheral (p) and normal skin area (n).
⋆: Difference (i.e. max−min) of homogeneity of the GLCM (quantization level = 64, distance = 4 pixels) among the four directions 0°, 45°, 90°, and 135° in central tumor (ct).
⋄: Skewness (κ) of saturation (S) is calculated for each of the 4 angle-wise subregions in central tumor (ct). Standard deviation (σ) of these skewness values is used as the feature.

The features are written in the "category : detail" format.

- Table II summarizes the result of the classification performance under the 10-fold cross validation test.

TABLE II
CLASSIFICATION PERFORMANCE OF THE THREE MODELS UNDER THE CONDITION OF %M > 90%

Model	#features ⋆		AUC †	%M	%N	%B	%S	min(%N, %B, %S)
Layered Model	20	(9, 6, 5)	**0.824**	**90.48**	**75.58**	**86.96**	**76.53**	**75.58**
Flat Model I			0.775	90.48	69.94	72.46	71.43	69.94
Flat Model II		(4, 4, 9, 3)	0.750	90.48	69.08	69.57	68.37	68.37
Layered Model	25	(9, 10, 6)	**0.856**	**90.48**	**82.51**	**82.61**	**80.61**	**80.61**
Flat Model I			0.802	90.48	74.57	75.36	74.49	74.49
Flat model II		(6, 3, 8, 8)	0.795	90.48	74.42	75.36	74.49	74.42
Layered Model	30	(11, 12, 7)	**0.864**	**90.48**	**82.66**	**84.06**	**82.65**	**82.65**
Flat Model I			0.821	90.48	75.58	76.81	75.51	75.51
Flat Model II		(8, 5, 9, 8)	0.802	90.48	76.59	76.81	78.37	76.59

†: AUC denotes the area of the receiver-operating characteristic (ROC) curve between %M and min(%N, %B, %S).
⋆: The numbers in the parenthesis denote those of the features assigned to each classifier in the order of the "MN-BS", "M-N", and "B-S" for the layered model and the "M", "N", "B", and "S" for the flat model II, respectively. The flat model I has all classifiers share the same 20, 25, or 30 features.

- The AUC is the area under the receiver-operating character-istic (ROC) curve. Fig. 5 shows the ROC curves drawn from the classification results by the layered model and the two flat models with 25 features.

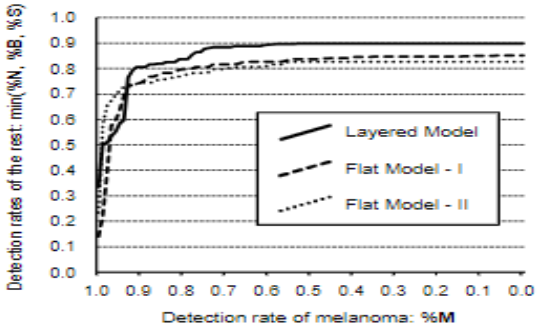

Fig. 5. ROC curves by the three models with 25 features.

The horizontal axis is the detection rate of melanoma (%M) which we define as the ratio of the correctly classified melanoma images over all

the melanoma images in the dataset (105 as described in Section II). The vertical axis is the minimum of the detection rates of nevus, BCC, and SK (min(%N, %B, %S)). They made the curves by optimizing the thresholds and the scaling factors of the linear classifiers to maximize min(%N, %B, %S) under the condition imposed on %M ranging to 100% from 0%.

- The reason for using the minimum is to measure the detection rate applicable to all three non-melanoma types of skin lesions.
- Larger area under the curve (AUC) indicates better performance.
- Seeing the ROC curves, they notice that min(%N, %B, %S) does not reach 100% even when %M is set to zero unlike the typical ROC curves seen in the studies of binary classification between melanoma and the rest. This is because min(%N, %B, %S) would reach 100% only if all the nevi, BCCs, and SKs could be perfectly classified as such, whereas in the binary classification, the detection rate of non-melanoma, i.e. SP (specificity) reaches 100% simply by increasing the threshold enough to dismiss all the melanomas as non-melanoma. This is the reason why the AUC in Table II seems comparatively lower than those reported in other conventional works.
- Finally, the %M, %N, %B, and %S in the table show the result under the condition that the detection rate of melanoma (%M) should be at least 90%.

6 Discussion

6.1 Classification of Cancers

- They impose additional conditions on detection rate of either 90% or 85% for BCCs (%B) whilekeeping that of 90% for melanomas (%M).
- Under such conditions, they measured the classification performance of the three models, table VI summarizes the results.

TABLE III
CONFUSION MATRIX (%) BY THE LAYERED MODEL WITH 25 FEATURES

result / dataset	Melanoma	Nevus	BCC	SK
Melanoma	**90.48**	6.67	0.95	1.90
Nevus	15.17	**82.51**	0.72	1.59
BCC	5.80	1.45	**82.61**	10.14
SK	5.10	7.14	7.14	**80.61**

The number of the features was set to 25 for the two flat models as well as the layered model shown in TableII. Thethreshold values and the scaling factors of the linear classifiers were adjusted in the following manner:

1. Keep %M and %B greater than the predefined values, i.e. shown in Condition columns in Table VI

TABLE IV
CONFUSION MATRIX (%) BY THE FLAT MODEL I WITH 25 FEATURES

result / dataset	Melanoma	Nevus	BCC	SK
Melanoma	**90.48**	6.67	2.86	0.00
Nevus	16.18	**74.57**	2.89	6.36
BCC	13.04	2.90	**75.36**	8.70
SK	12.24	4.08	9.18	**74.49**

2. Maximize the minimum of the two other detection rates (min(%N,%S))

- Note that "unattainable" in the table indicates that no threshold value and scaling factor for the flat model I meets these conditions.
- They saw that the layered model showed much higher performance than the two flat models.

- Compared to the test without any condition on %B (features = 25 in Table II), the detection rates of nevus and SK (%N, %S) decreased.
- This is an inevitable trade-off for achieving high detection rates for both melanomas (%M) and BCCs (%B).

  They think that the appropriate adjustment is to give the first priority to the detection rate of melanoma (%M) then the second priority to that of BCC (%B) while keeping those of nevus and SK (%N,%S) within an acceptable range.
- Besides, they also tested a cascademodel based on the notion that SK looks more similar to MSLs than BCC.
- This model firstly distinguishes "MSLs and SK" from BCC, then MSLs from SK, and finally melanoma from nevus. However, it was even inferior to both of the flat models.

6.1.1 Feature Interpretation

- For more insight into the developed model, they examined how the selected featurescontributedto the classification.
- Fig. 6 shows examples of images classified correctly by the layered model with 25 features.

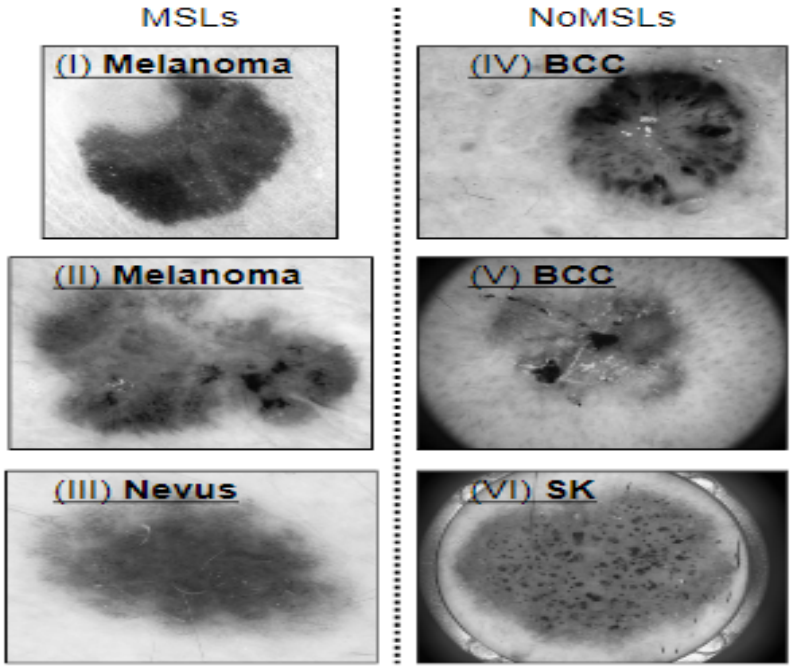

Fig. 6. Examples of correctly classified images.

The left-column shows MSL images: (I) and (II) are melanomas, and (III) is a nevus. The right-column shows NoMSL images: (IV) and (V) are BCCs, and (VI) is a SK.

- Fig. 7 shows the scatter plot of the first feature "Col: σ (V: p)" and the second feature "Sub:$\sigma(\mu$(R: ct)) [angle-16]" of the classifier "M-N" (see Table I).

  The circles and dots represent the images of melanomas and those of nevi, respectively.The plots corresponding to the images (I), (II), and (III) in Fig. 6 are specified by the arrows. The dashed line shows the classification boundary for the 90% melanoma detection rate. This is a rough criterion to distinguish between the two types of skin lesions.

- Although some images are still misclassified due to the overlapped area, they used more features to improve the classification performance as specified by features in Table II.

  1. The first feature "Col: σ (V: p)" is the standard deviation of luminance in the peripheral. This feature tends to be higher for melanomas than nevi.
     For example, the nevus image (III) in Fig. 6 shows a gradual increase of luminance going outward from the center while the melanoma image (I) shows an irregular decrease of luminance at the peripheral, causing a relatively high contrast against the surrounding normal skin area.
  2. The second feature "Sub:$\sigma(\mu$(R: ct)) [angle-16]" is the difference of the red channel between the angle-wise subregions shown in Fig. 2. This feature was also larger for melanomas than nevi possibly because melanomas tend to have an uneven or irregular color distribution as established by the ABCD-rule and the 7-point check list, two common references for melanoma diagnosis.

- Regarding NoMSLs, Fig. 8 shows the scatter plot of the first and the second features of the classifier "B-S".

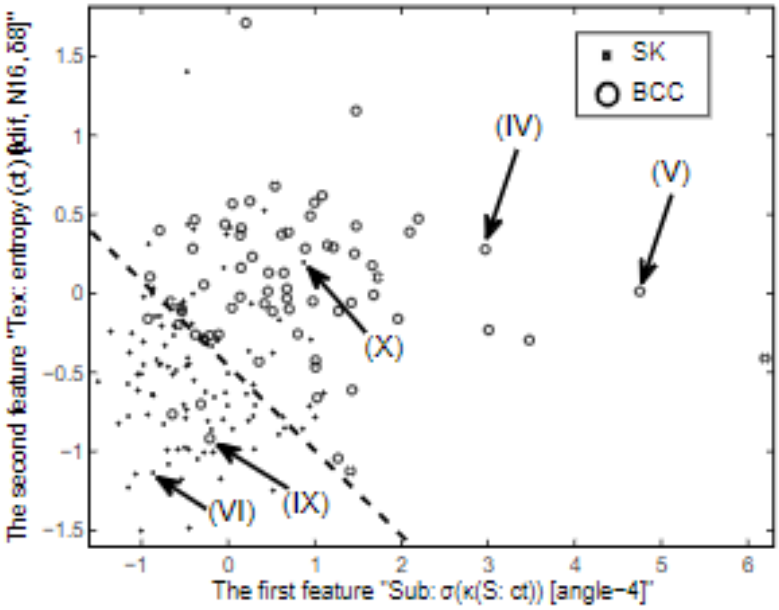

Fig. 8. Plots of the two selected features of the "B-S".

The plots corresponding to the images (IV), (V), and (VI) in Fig. 6 are also shown.
The dashed line shows the boundary for the BCC detection rate of 90% based on the two features.

1. The first feature "Sub:$\sigma(\mu$(S: ct)) [angle-4]" is the differ-ence in skewness on saturation channel between the angle-wise subregions. This feature was higher for BCCs than SKs on average mainly due to the presence of different local objects seen specifically in BCCs such as dark pigments (IV) and blood vessels (V).
2. The second feature "Tex: entropy (ct) [dif, N16, 8]" is the directionality of the coarseness in the central tumor. Some of the SKs in our datasets showed a lot of holes with no preference of direction or location as seen in (VI), making this feature especially low.

- Fig. 9 shows the scatter plot of the first and the second features of the classifier "MN-BS".

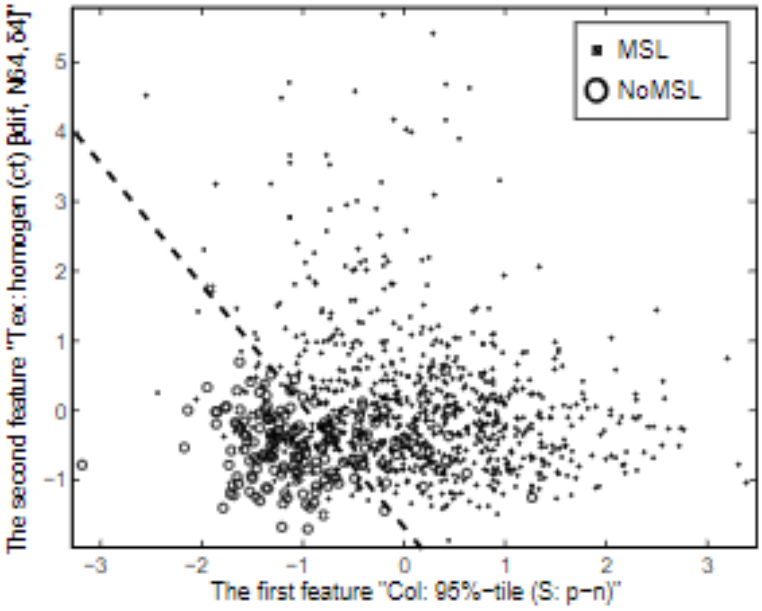

Fig. 9. Plots of the two selected features of the "MN-BS".

The dashed line is the boundary for 90% classification rate of MSLs. Looking a the distribution, this classification is more accurate than that between melanoma and nevus shown in Fig. 7. The AUC computed on this scatter plot was 0.911, which is higher than 0.888 on Fig. 7 and 0.891 on Fig. 8

- Misclassifications still occurred due to the varied appearances of the skin lesions.

- Fig. 10 shows examples of images misclassified by the layered model.

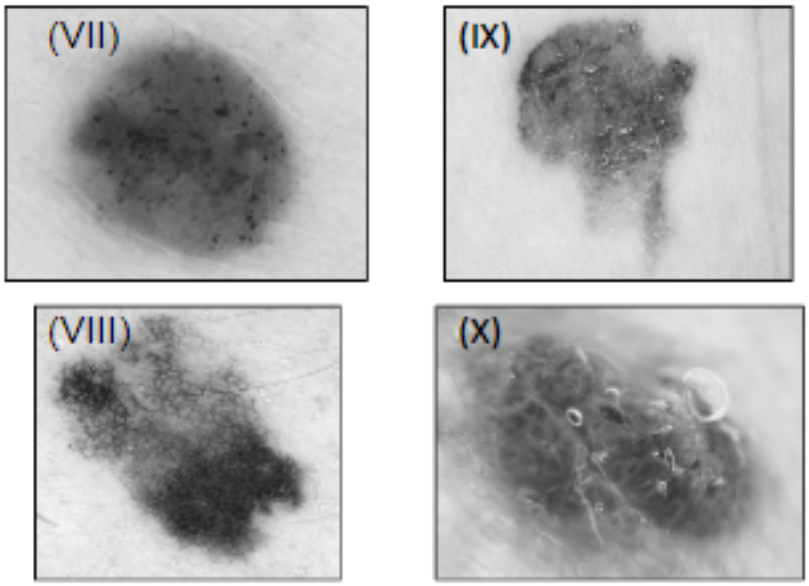

Fig. 10. Examples of misclassified images. (VII) melanoma misclassified as nevus. (VIII) nevus misclassified as melanoma. (IX) BCC misclassified as SK. (X) SK misclassified as BCC.

These images are also specified by the arrows in Figs. 7 and 8.

- The melanoma (VII) was misclassified as nevus due to the lack of certain characteristics of melanoma, e.g. the difference in color between the angle-wise subregions. Nevertheless, the presence of the irregular dark globules might be seen as a sign of melanoma. Dealing with such detailed patterns will be the goal of future work.

- Despite these difficulties, the layered model with 25 features achieved a detection rate of 90% for melanomas and over 80% for nevi, BCCs, and SKs. These results might seem inferior to those reported in other studies of the binary classification between melanoma and nevus such as 93.3% SE and 92.3% SP.

- However, this seems inevitable since our methods distinguished among four types of skin lesions instead of two.

- Our analysis used a comparatively large number of images from different data sources, which makes the analysis more realistic.

- Note that the performance of the method will possibly improve by using non-linear classifiers, e.g. SVM.

- In this study, they dealt with four types of skin lesions, while they did not include squamous cell carcinoma (SCC) the second most common skin cancer due to unavailability of datasets. they are planning to deal with those skin lesions in near future.

7 Conclusion

- In this paper, they proposed a method to distinguish among melanomas, nevi, BCCs, and SKs.
- For the classification model, they introduced a layered model for task decomposition and two flat models to serve as the baseline.
- They evaluated the models with 964 dermoscopy images and showed that the layered model outperformed the two flat models.
- The layered model with 25 features achieved a detection rate of 90% for melanomas and over 80% for each of the three other types of skin lesions. The result of this study shows promise for broadening the range of users for classification and enhancing the capability of computer-aided skin lesion classification.

8 Acknowledgment

- This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2014.2348323, IEEE Transactions on Biomedical Engineering

- IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 1, NO. 1, JANUARY 2014

- Manuscript received March 25, 2014; revised July 13, 2014; accepted August 3, 2014. Date of publication XXX; date of current version xxx.

- This research was supported in part by Japan Science and Technology A-STEP Program (AS251Z01435P, 2013), NIH, and ACS. Asterisk indicates corresponding author.

- *K. Shimizu and H. Iyatomi are with Department of Applied Informatics, Hosei University, Tokyo, Japan (e-mail: kohei.shimizu.66@adm.hosei.ac.jp; iyatomi@hosei.ac.jp).

- M. E. Celebi is with Department of Computer Science, Louisiana State University in Shreveport, LA, USA (e-mail: emrecelebi1980@gmail.com).

- K.A. Norton is with Departmentof Biomedical Engineering, Johns Hopkins University, MD, USA (e-mail: kerri.norton@gmail.com).

- M. Tanaka is with Department of Dermatology, Tokyo Women's Med-ical University, Medical Center East, Tokyo, Japan
(e-mail: masaru-tanaka@1984.jukuin.keio.ac.jp).

- Digital Object Identifier 10.1109/TBME.2014.XXXXXXX

9 Summary

9.1 Motivation

Most of the conventional works handled only melanocytic skin lesions (MSLs) such as melanomas and nevi which originate from melanocytes, whereas non-melanocytic skin lesions, (NoMSLs) indicating all the other pigmented skin lesions except MSLs such as basal cell carcinomas (BCCs) and seborrheic keratoses (SKs) have been relatively neglected. This is because melanoma is the most fatal skin cancer and especially difficult to differentiate from nevus even by expert dermatologists. However, BCC is also harmful and accounts for 80% of the skin cancer incidences. SKs are observed in most Caucasian people above the age of 50 and are sometimes confused with melanomas.

Although classification of NoMSLs is considered to be easier than that of MSLs for expert dermatologists, it is not always easy for inexperienced dermatologists or physicians with different expertise. Therefore, if they open the system also for those potential users, as was the case with the abovementioned internet-based system, it is necessary to handle both MSLs and NoMSLs.

9.2 Dataset

- Melanoma: 105 images (30 from Keio University Hos-pital and 75 from University of Naples and Graz), a malignant melanocytic tumor (MSL), the most fatal skin cancer.
- Nevus: 692 images (448 from Keio University Hospital and 244 from University of Naples and Graz), a benign melanocytic tumor (MSL), often difficult to differentiate from melanomas.
- Basal cell carcinoma (BCC): 69 images (20 from Keio University Hospital and 49 from Tokyo Women's Med-ical University), a malignant non-melanocytic tumor (NoMSL), the most common skin cancer.
- Seborrheic keratosis (SK): 98 images (42 from Keio University Hospital and 56 from Tokyo Women's Medical University), a benign non-melanocytic tumor (NoMSL), which commonly occurs in the elderly and is sometimes confused with melanomas.

9.3 Framework

- These methods usually consist of three steps:
    1. Border detection of skin tumor
    2. Feature extraction
    3. Classification
- The border detection process finds the border of the tumor in the dermoscopy image, which is essential for accurate skin lesion classification.

- Several methods have been proposed such as dermatologist-like method, SRM, hybrid thresholding, threshold fusion, and so on.
- The feature extraction process obtains discriminating image features that facilitate classification such as general color statistics, contour shape, and texture information.
- Wavelet coefficients that capture color and shape information have also been investigated.
- The classification process determines the type of skin lesions from the extracted image features.
- General pattern classifiers such as linear discriminant classifier, k-NN, artificial neural networks, and SVMs are often used.
- Based on the aforementioned three steps, researchers have improved the automated classification methods. Although there are several limitations, these studies reported superior classification performance compared to experts.
- They developed an internet-based melanoma screening system (current URL is https://dermoscopy.k.hosei.ac.jp) which they continually update to improve accuracy and reliability.
- The abovementioned conventional studies have several problems:
    1. Only limited types of skin lesions are acceptable for classification
    2. The systems do not explain the reasons for the classification results
    3. The systems were developed and evaluated with only ideal condition images and did not consider the condition of test images
- They developed a general border detection algorithm for MSLs and NoMSLs. Finding the border of NoMSLs was a challenging task because they often have unclear borders. With this sophisticated algorithm, they found that a linear classifier with only two image features ('skewness of bright region on the major axis' and 'difference in blue intensity between the peripheral and the normal skin') discriminated MSLs from NoMSLs with performance of 98.0% SE and 86.6% SP.
- They further developed a system to detect melanomas from other MSLs (nevi) and NoMSLs. Using 548 MSL and 110 NoMSL images, the system achieved 97.6% SE and 87.7% SP (89.5% SP for nevi and 79.1% SP for NoMSLs).
- However, this study focused only on discriminating melanomas from all other lesions, thus clumping BCCs with benign skin tumors.
- In this paper, they propose a method to distinguish among four types of skin lesions: melanoma, nevus, BCC, and SK, using a significantly larger dataset.

9.4 Challenge

9.4.1 Solved Challenges

1. Finding the border of NoMSLs was a challenging task because they often have unclear borders. With this sophisticated algorithm, they found that a linear classifier with only two image features ('skewness of bright region on the major axis' and 'difference in blue intensity between the peripheral and the normal skin') discriminated MSLs from NoMSLs with performance of 98.0% SE and 86.6% SP.

2. They resized the images such that the major axis of the whole tumor was 512 pixels in length due to the disparate image resolutions and to reduce computation time.

3. For the flat model I, it is necessary to take all four classifiers into consideration when selecting a feature to add or remove in the stepwise method. This issue was addressed in which examined multiclass classification. It suggests two methods: either to optimize 'average' or 'maximum' of the four error reduction amounts associated with the four outputs. They chose the 'average' method because it showed better performance in our preliminary experiments.

9.4.2 Unsolved Challenges

1. The melanoma (VII) was misclassified as nevus due to the lack of certain characteristics of melanoma, e.g. the difference in color between the anglewise subregions. Nevertheless, the presence of the irregular dark globules might be seen as a sign of melanoma. Dealing with such detailed patterns will be the goal of future work.

2. In this study, they dealt with four types of skin lesions, while they did not include squamous cell carcinoma (SCC) the second most common skin cancer due to unavailability of datasets. they are planning to deal with those skin lesions in near future.

9.5 Results

The layered model with 25 features achieved a detection rate of 90% for melanomas and over 80% for each of the three other types of skin lesions. The result of this study shows promise for broadening the range of users for classification and enhancing the capability of computer-aided skin lesion classification.