

Clinical Skin Lesion Diagnosis using Representations Inspired by Dermatologist Criteria

Jufeng Yang[†], Xiaoxiao Sun[†], Jie Liang[†], Paul L. Rosin[‡]

[†]College of Computer and Control Engineering, Nankai University, Tianjin, China

[‡]School of Computer Science and Informatics, Cardiff University, Cardiff, UK

yangjufeng@nankai.edu.cn, {sunxiaoxiao, liang27jie}@163.com, RosinPL@cardiff.ac.uk

Abstract

The skin is the largest organ in human body. Around 30%-70% of individuals worldwide have skin related health problems, for whom effective and efficient diagnosis is necessary. Recently, computer aided diagnosis (CAD) systems have been successfully applied to the recognition of skin cancers in dermoscopic images. However, little work has concentrated on the commonly encountered skin diseases in clinical images captured by easily-accessed cameras or mobile phones. Meanwhile, for a CAD system, the representations of skin lesions are required to be understandable for dermatologists so that the predictions are convincing. To address this problem, we present effective representations inspired by the accepted dermatological criteria for diagnosing clinical skin lesions. We demonstrate that the dermatological criteria are highly correlated with measurable visual components. Accordingly, we design six medical representations considering different criteria for the recognition of skin lesions, and construct a diagnosis system for clinical skin disease images. Experimental results show that the proposed medical representations can not only capture the manifestations of skin lesions effectively, and consistently with the dermatological criteria, but also improve the prediction performance with respect to the state-of-the-art methods based on uninterpretable features.

1. Introduction

The skin is directly exposed to the air, which leads to skin disease being one of the most common human illnesses. It pervades all cultures, occurs at all ages, and affects the health of 30% to 70% of individuals, with even higher rates for the at-risk subpopulations [21, 22]. A common technique used by dermatologists for diagnosing skin diseases is the dermoscope which enables observation of the latent structures of skin lesions, *i.e.*, a region suffering from disease, whose effects are otherwise invisible to the

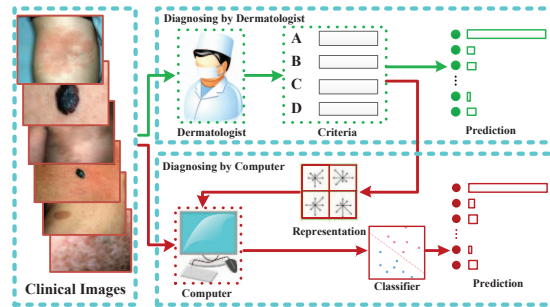


Figure 1. Given a patient with a common skin disease (the left box), dermatologists make a diagnosis by observing the appearance of the lesions (the top right box). This paper aims to design a skin disease recognition system (the bottom right box) based on clinical images and dermatological criteria. Firstly, the medical information observed by the doctors during diagnosis is exploited. Then, measurable medical representations for skin lesions are designed for diagnosis.

naked eye. Dermoscopy is usually used for melanomas and other kinds of skin cancers with pigmentation [13]. However, it is difficult to access a dermoscope in resource-poor regions [20], and moreover it is unnecessary for most of the common skin diseases. Therefore, developing an effective skin disease diagnosis system based on easily accessed clinical images would be beneficial and could provide low-cost, universal access to more people.

Experienced dermatologists diagnose skin diseases based on comprehensive medical criteria which have been verified to be useful, *e.g.*, the ABCD rule [1, 32] and the 7-point checklist [3], *etc.* There are also specialized criteria for individual aspects [38] which focus on the diagnosis of a specific kind of skin disease, *e.g.*, the structure difference [12].

Computer aided diagnosis (CAD) systems based on dermoscopic images and medical criteria have been developed over the past decade [4, 25], yet few of them attempt to diagnose the lesions of clinical skin disease images. Moreover, dermoscopy based diagnosis is limited by the enor-

mous expenses and the inconvenience access [20,35], while experienced dermatologists are able to make effective diagnosis for most common diseases by simply observing the lesions with the naked eye. There are three primary difficulties for recognizing skin disease from a clinical image captured by easily-accessed devices, *e.g.*, smart-phone, as compared to dermoscopic images. (1) The clinical images are captured with various illumination conditions. (2) The lesions in clinical images have non-uniform focal lengths and the size of viewing frame is not consistent. (3) There are far more categories of diseases in clinical images than dermoscopic images, since dermoscopy is designed to diagnose mainly skin cancers, *e.g.*, melanomas. Despite the existing difficulties, an automatic diagnosis system for clinical images would be helpful to junior doctors and regular patients.

There exist several works on diagnosing skin disease by analyzing digital images [11, 29, 33]. While most of them employ label information that has been provided by expert doctors to design the diagnosis systems, they generally do not consider medical meaning during the diagnosis process [11, 33]. As a result, the representation they use is not interpretable which means it is less helpful for convincing the patients. Meanwhile, unlike other recognition problems, a diagnosis system needs to be accepted by doctors, and therefore needs to be consistent with evidence-based medicine [9].

In this paper, we propose to exploit the dermatologists' criteria to develop effective and interpretable visual representations for lesions in clinical images, of which the main idea is illustrated in Figure 1. To incorporate the common criteria into medical representations, we first investigate and verify the mapping relationship between both concepts. Accordingly, we design representations for skin lesions relating to three aspects, *i.e.*, structure, color and shape of the lesions. First, the **structure** of each skin disease is represented by the distribution of the texture. It is measured mainly by the symmetry property of texture in multiple spaces. Second, for the representation of **color**, we preserve only the colors associated with skin lesions for effective representation from ColorName [39, 40]. We also introduce continuous values for each color to distinguish between diseases with the same color name but different degrees of color, *e.g.*, pityriasis rosea and lichen planus. Third, to represent the **shape** of lesions, we consider the peripheral symmetry and the constrained compactness, *etc.* We introduce an approximate measurement for the area of the lesion considering the associated swollen or congested region.

The contributions of this paper are summarized as follows. Firstly, we verify the measurability of the criteria employed by the dermatologist, which provides the theoretical basis to design the diagnosis system for skin disease. Then, we propose comprehensive medical representations for skin

lesions in clinical images according to the dermatological criteria. Finally, based on the representations, we develop a clinically oriented diagnosis system which can identify skin diseases from clinical skin lesion images. Extensive experiments on the SD-198 dataset [33] demonstrate the effectiveness of the proposed diagnosis system.

2. Related Work

We review the previous works from two aspects, *i.e.*, the diagnosis of skin diseases by dermatologists and the computer aided diagnosis (CAD) system for skin lesions.

2.1. Diagnosing Skin Diseases by Dermatologists

Common skin diseases are primarily diagnosed by observation. In the diagnosis process, doctors make the diagnosis and provide suggestions for treatment according to the visual information (sometimes along with a questionnaire) and the empirical criteria. There are few indistinguishable diseases which require dermoscopic analysis, biopsy or histopathological examinations, *etc.* [17, 24]. In the medical community, much research focuses on the diagnosis of skin disease based on the characteristics of the lesions, *e.g.*, color [15], shape, *etc.* For example, color information can help with the diagnosis and the recommendations for treatment of acne vulgaris for patients [34]. In addition, atopic dermatitis is diagnosed based on the morphological appearance of the lesions [37]. For most of the pigmented skin lesions and other diseases, the ABCD criteria [32] and the 7-point rule [3] are usually employed for diagnosis. As illustrated in [27], faced with a patient or a clinical image, the dermatologists analyze the information they observed from the skin lesions and incorporate the empirical criteria [32] to make the diagnosis. However the diagnosis of skin lesions by observation is always subjective [30] and relies on the vast experience of dermatologists. Therefore, the guidance or suggestions from a decision support system can help the dermatologists to make a more objective and confident diagnosis than that based only on individual observation and experience. The challenge of designing a successful CAD system lies on the incorporation of dermatological criteria with medical meaning, which is developed in this paper.

2.2. CAD for Skin Lesions

Earlier research on CAD systems for skin lesions mainly relates to pigmented skin disease (*e.g.*, melanoma) detection [8] and classification [16], of which the detected or predicted results do not satisfactorily improve the diagnosis of the clinicians [25]. Recently, with the development of computer vision and artificial intelligence, there arose many explorations of using computers to analyze skin lesions. [31] discussed the potential benefits of applying digital imaging to dermatology. [19] proposed an approach to

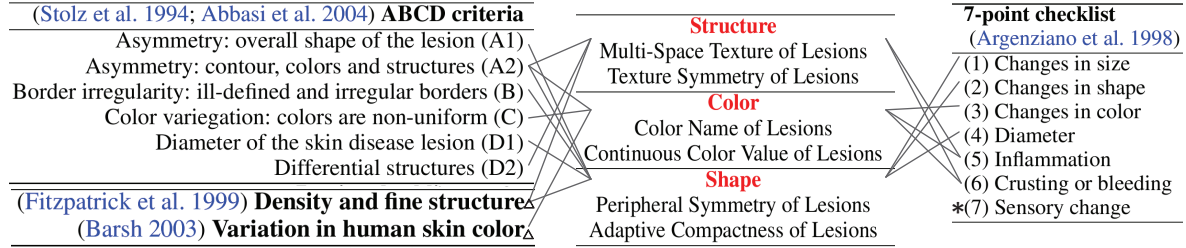


Figure 3. The mapping to visual components from dermatological criteria, *i.e.*, ABCD, 7-point, and the two specialized criteria which focus on a single aspect. A1, D1 and A2, D2 are related to different versions of the ABCD criteria as described in the section on Criteria. “*” indicates that the criterion corresponds to the physical sensation of patients which cannot be measured in lesion images, *e.g.*, numbness and itching. The criteria with “ Δ ” focus on one aspect, *e.g.*, structure or color of the lesions. As shown, the dermatological criteria are highly correlated with the three visual components.

The ABCD criteria are widely used for diagnosing pigmented skin lesions in both clinical images and dermoscopic images. Although there are various combinations of which the meanings are different for clinical images and dermoscopic images, they contain the same aspects of skin lesion, *i.e.*, structure, color and shape (asymmetry, border), *etc.* The Glasgow 7-point checklist [3] is also a set of criteria. It contains three major aspects, *i.e.*, change in size, shape and color, and four minor aspects, *i.e.*, diameter, inflammation, crusting or bleeding and sensory change, among which inflammation and crusting or bleeding are considered as properties of the structures (see Figure 2 (11, 13)).

Rather than considering multiple criteria, much research concentrates on a single aspect of the skin disease, *e.g.*, the structure difference [12], the color variation [5, 36], the shape contrast [23], *etc.* These criteria can guide dermatologists to make a decision by observing the specific properties of skin lesions. Therefore, in this paper, we consider the comprehensive attributes of the skin diseases by combining the researches concentrated on each criterion and the original multiple criteria, *e.g.*, ABCD or 7-point, the details of which can be found in the next section.

4. Proposed Medical Representations

The bottom row of Figure 2 shows the initial measurable representations derived by the computer from the corresponding clinical images, *e.g.*, Figure 2 (4) is the border of the lesions in Figure 2 (3). As shown, the characteristics presented in skin lesions can be distinguished not only by dermatologists using the naked eye according to the criteria introduced in the last section, but can also be represented by the computer. Specifically, the medical representation of the aforementioned criteria can be generated as follows. The area, the diameter (D) and the border (B) of lesions are determined by the number of pixels and the principal axis of the connected region in computers (see Figure 2 (2, 4)), where Asymmetry (A) can also be calculated by the geometric information of the connected region. For the cri-

terion of color (C), the computer measures it in different color spaces (see Figure 2 (6, 8 and 10)), which enables discrimination even for skin diseases with subtle changes of color, *e.g.*, the tinea in Figure 2 (9). In addition, the clinical manifestations of the skin lesions, *e.g.*, the inflammation, crusting or bleeding in the 7-point checklist [3], can be quantitatively determined by the change of texture, color and structure, on the surface of the skin (see Figure 2 (12, 14)). As shown in Figure 3, the aforementioned criteria are highly correlated to three visual components, *i.e.*, structure, color and shape. For example, the Asymmetry property in the ABCD rules, the “changes in size” and the “changes in shape” properties in the 7-point checklist can be considered as the different aspects for the visual component of the shape. We design representations for each visual component based on an interpretable way of which the details are illustrated as follows.

4.1. Structure Representation

The lesions of different diseases have various epidermal structures, *e.g.*, scales, lumps, scabs, hemorrhage, which can be measured by the textures in the clinical images. Meanwhile, the structure of skin disease has different distributions, *e.g.*, single and homogeneous lesions or structureless, which can be measured by the symmetry of the structure.

4.1.1 Multi-Space Texture of Lesion (MST-L)

While the dermatologist can easily observe the skin lesions of a patient, the clinical images of the skin disease are captured under different illumination conditions which can affect the calculation of the representations. To represent the structure of the lesions effectively, we calculate the texture representation based on different color channels which reduces the influence of the environment. We propose the multi-space texture of lesions $MST(x)$ for each clinical image x as follows:

$$MST(x) = [G_i(x)]_{i=1}^K, \quad (1)$$

where $G_i(x)$ is the set of the texture features extracted from the i -th color channels and K denotes the number of spaces. we employ three channels in our work, *i.e.*, Hue, Saturation and Brightness, from each of which we extract SIFT features for the representation.

4.1.2 Texture Symmetry of Lesion (TS-L)

The asymmetry of structures for skin lesions turns out to be an effective criterion. Therefore, we propose a representation based on the texture symmetry of lesions. Firstly, we divide the lesion region detected by MBD+ [41] into two parts, *i.e.*, $L(x)_1$ and $L(x)_2$, according to its principal axes.

Then, we extract the texture features for each of the parts. We finally represent the texture symmetry $S_i(x)$ of lesions in the i -th color channels of the image x as follows:

$$TS_i(x) = [G_i(L(x)_1), G_i(L(x)_2), S_i(x)]. \quad (2)$$

Here, we define the symmetry representation $S_i(x)$ as:

$$S_i(x) = \{|g_{ij}^1 - g_{ij}^2|\}_{j=1}^d, \quad (3)$$

where d is the dimension of the extracted features, g_{ij}^1 and g_{ij}^2 are the j -th entry of $G_i(L(x)_1)$ and $G_i(L(x)_2)$, respectively. We measure the texture symmetry in Hue space since it is scale- and shift-invariant with respect to the light intensity.

4.2. Color Representation

The color of the lesion is a crucial criterion for diagnosing skin diseases. For instance, regular dermatitis appears red, while stasis dermatitis is always dark-brown. In addition, due to variations in the illumination, an appropriate color normalization operation is helpful. In this section, we design two representations based on the color of the lesions.

4.2.1 Color Name of Lesion (CN-L)

Objects in the real world are colorful, but skin diseases are restricted to several colors [36]. In computer vision, the color category is modeled by a mapping function which assigns the three channel values of a pixel into a sparse space spanned by M colors. Inspired by [6], we calculate the probability vector $P = [p(C_l|c)]_{l=1}^M$ for each color bin in the $L * a * b$ space (which is sparsely sampled from the RGB-cube) as follows:

$$[p(C_l|c)]_{l=1}^M \propto \sum_i^N p(C_l|c_i) g^\sigma(|c_i - c|_{Lab}), \quad (4)$$

where c denotes the original value of the color bin, c_i is $L * a * b$ -value for c , $N = 387$ is the total number of the color bins and C is the set of basic colors used in our

work. According to [39] we set the weighting kernel g^σ as a Gaussian kernel with $\sigma = 5$. Based on the investigation of skin lesions in [5], we set $M = 8$ and $C = \{red, pink, purple, yellow, white, black, brown, blue\}$ in this work. The calculation process of $p(C_l|c)$ is similar to that in [39]. Finally, we define the color name of lesions $CN(x)$ as:

$$CN(x) = \operatorname{argmax}_{C_l} [p(C_l|c)]_{l=1}^M. \quad (5)$$

4.2.2 Continuous Color Values of Lesion (CCV-L)

Different from the ColorName feature which assigns a fixed vector for each color name, *e.g.*, for $C_l = red$: $value = (1, 0, 0)$, in this work we assign a continuous value for each lesion to indicate different degrees of the color. For each pixel in the image, we define the representation of continuous color values $CCV(c)$ for each color bin c as:

$$CCV(c) \propto p(C, c) \times \theta(c), \quad (6)$$

where $p(C, c)$ indicates the probability of mapping the color bin c into its nearest color name C . The $\theta(c)$ is the weighting value of the pixel which is defined as:

$$\theta(c) = \sum_{|c|} n(c)u(c), \quad (7)$$

where $n(c)$ is the frequency of the corresponding color in the image, $u(c)$ is the color value of c in the RGB space.

4.3. Shape Representation

The shape of skin lesions is an important cue to diagnose both the category and the degree of the skin disease. For example, a nevus lesion can be a circle while a lesion of psoriasis appears to be scattered. In this paper, we propose two representations according to the shape of the lesions: (1) the symmetry of the shape and (2) the constrained compactness of the lesions, both of which are calculated from the detected lesion regions.

4.3.1 Peripheral Symmetry of Lesion (PS-L)

We first detect the lesion region in each clinical image x using the MBD+ method. It is then divided into two parts, *i.e.*, $L(x)^1$ and $L(x)^2$, with approximately equal area and periphery based on the consistent characteristic of the skin lesions. Specifically, we find the longest straight line segment within the lesion, and then partition the region by the segment's perpendicular bisector. Finally, we represent the peripheral symmetry of lesions based on the complementarity of the two parts:

$$PS(x) = F(A(L(x)^1), A(L(x)^2)), \quad (8)$$

where $A(\cdot)$ denotes the extracted feature of lesions and $F(\cdot, \cdot)$ denotes the concatenation function applied to the two features.

4.3.2 Adaptive Compactness of Lesion (AC-L)

For a skin disease region, the degree of its approximation by a circle can be useful for diagnosis, and can be represented by the criterion of compactness which is defined as:

$$\text{Com} = \frac{4\pi A}{P^2}, \quad (9)$$

where A denotes the area and P denotes the perimeter of the lesion. However, the lesions of skin disease can be surrounded by swollen or subcutaneous hemorrhage which may also be detected in clinical images. Note that those regions always appear as lighter colors than those in the center of lesions. Therefore, we adaptively calculate the area of lesions with dynamic weights according to the pixel importance when measuring the compactness. We define the area A_L of the lesions as follows:

$$A_L = \sum_{z \in L(x)} p(C|c, z), \quad (10)$$

where z denotes the pixel in the lesion $L(x)$. $p(C|c, z)$ is the probability of mapping color to a specific color category in the Color Name feature, which reflects the significance of the pixel z being the center of the lesion.

5. Experiments

5.1. Setup

Dataset. The SD-198 dataset [33] is the largest publicly available dataset in this field, containing 198 skin diseases and 6,584 clinical images, which are collected by digital cameras or mobile phones. The images vary in color, exposure, illumination and scale, and include a wide range of patients with different ages, gender, locations of disease, colors of skin and different stages of the disease. We use the standard train/test split provided by SD-198, which has 3,292 training and 3,292 testing images.

Evaluated Representations. We compare the proposed representations with several standard low-level features from three visual components, *i.e.*, texture (SIFT, HOG, LBP, BRIEF, SURF, Wavelet and ORB), color (Color Histogram (CH), Color Name (CN) and ColorSIFT), Border (GIST, Gabor, Prewitt, Sobel and Canny). Meanwhile, we visualize the different kinds of designed representation in Figure 5. We then fuse the best features of each component (as determined by experiments) for a comprehensive combination. Meanwhile, we compare the proposed representations against the features derived by state-of-the-art

convolutional neural networks (CNNs), including CaffeNet, VGGNet and GoogleNet. In the experiments, we also extract features from the fine-tuned models in order to provide a fair comparison.

Metrics. In this work, we generate the recognition results of the contrasted representations using classifiers including k nearest neighbors (KNN), support vector machine (SVM) and random forest (RF) (the parameters of classifiers can be found in supplementary material). We employ classification accuracy (ACC) to evaluate the performance of the tested methods. Considering the recall rate for skin disease recognition, we also evaluate sensitivity

$$\text{SE} = \frac{1}{q} \sum_{i=1}^q \frac{(TP)_i}{(TP)_i + (FN)_i}, \quad (11)$$

where q denotes the number of classes, $(TP)_i$ and $(FN)_i$ denotes the number of true positive and false negative estimations for each class, respectively. The SE calculates the probability of being diagnosed as positive (*i.e.*, the percentage of sick people who are correctly identified as having the condition) [2]. Larger values of both ACC and SE reflect better diagnostic performance.

5.2. Effectiveness of the Visual Representations

We report the diagnosis performance of different representations as well as their combinations in Table 1. As shown, classifiers with a single low-level representation (#1~15) do not perform well. The best features of the low-level components barely reach about 31% ACC (SURF with SVM) and 29% SE (CN with RF), mainly because the commonly used features can neither reflect the comprehensive medical information relating to the lesions nor eliminate the noise in clinical images. Although the integration of components represented by low-level features (#16) improves the recognition performance, the best result is only 48.06% for ACC with RF. However, these results still suggest that the pathological changes of skin lesions can be captured by the visual components, which is consistent with Figure 2.

All the proposed representations with the inclusion of accepted dermatologist criteria improve diagnostic performance. For example, CCV-L (#20, which is the extension of the CN in #9 with same dimensionality) measures the lesions by mapping the region to a continuous color space spanned by eight color categories since the lesions with same color name yet different color degrees may correspond to different skin diseases. It achieves an accuracy of 45.32% and a sensitivity of 45.70% with RF, both of which show around 20% improvement against the original CN feature. The integration of the components represented by the proposed representations (#23) further improves the diagnostic performance by more than 10% compared to the integration of the components represented by low level features. It

Table 1. Diagnosis performance on the SD-198 dataset with different representations and different classifiers, *i.e.*, KNN, SVM and RF. CH (#8) denotes the Color Histogram feature and CN (#9) denotes the Color Name feature. As shown, the proposed representations (#17~22) show better performance than the low-level features (#1~15). Also, the integration of the components (#23) are more discriminative against the integration of components represented by low-level features (#16).

	Components	#	Features	Dimension	KNN		SVM		RF	
					ACC	SE	ACC	SE	ACC	SE
Baseline	Texture	1	SIFT	21000	20.35	19.17	25.55	24.75	21.42	21.25
		2	HOG	12400	19.14	17.85	17.62	14.45	10.54	10.66
		3	LBP	23200	15.13	14.80	18.89	14.69	14.61	13.24
		4	BRIEF	19200	16.74	15.62	12.21	8.39	15.67	15.03
		5	SURF	38400	17.47	16.50	31.17	25.35	27.34	26.52
		6	Wavelet	256	15.94	15.52	14.82	12.73	13.37	14.02
		7	ORB	19200	20.53	21.44	23.21	22.94	18.86	17.46
	Color	8	CH	256	12.33	12.58	4.19	4.41	18.77	16.81
		9	CN	21000	20.02	20.10	20.23	21.62	27.64	28.73
		10	ColorSIFT	21000	21.29	19.62	22.51	21.43	28.49	27.24
	Border	11	GIST	512	21.93	21.52	16.49	17.19	15.01	12.33
		12	Gabor	4000	13.67	13.00	10.15	8.62	13.73	12.43
		13	Prewitt	900	12.55	13.14	11.91	10.76	11.27	10.87
		14	Sobel	10000	12.27	12.03	10.42	10.18	13.46	12.46
		15	Canny	10000	15.22	17.16	13.91	14.51	16.46	15.20
	Integration	16	1&10&11	2500	47.36	47.23	46.84	47.24	48.06	46.73
Ours	Structure	17	MST-L	21000	44.99	45.62	48.06	46.38	43.23	42.73
		18	TS-L	21000	47.30	47.80	48.94	47.21	43.92	43.07
	Color	19	CN-L	21000	42.50	43.24	38.91	39.78	44.59	46.21
		20	CCV-L	21000	42.80	43.97	40.13	39.22	45.32	45.70
	Shape	21	PS-L	10000	30.04	30.47	38.58	38.29	38.94	36.87
		22	AC-L	10000	31.50	29.75	39.73	38.92	37.61	35.42
	Integration	23	18&20&22	3000	57.62	56.41	56.47	53.15	57.81	56.65

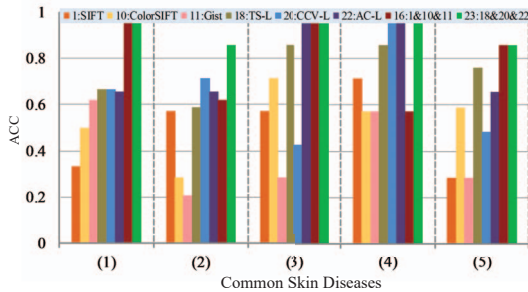


Figure 4. The detailed performance comparison of classification accuracy (ACC) over six kinds of commonly encountered skin diseases, *i.e.*, (1) Acne Vulgaris, (2) Acute Eczema, (3) Favre Racouchot Nail Ridging, (4) Pomade Acne, (5) Syringoma in SD-198 dataset with different representations. The representations in this figure correspond to the rows in Table 1, *e.g.*, the last column in each group employ the #23 representation of Table 1.

verifies that the three visual components corresponding to different medical criteria are complementary to each other, thus their incorporation provides a convincing diagnosis. Note that the diagnostic performance of both ACC and SE in Table 1 varies only slightly when different classifiers are used, but substantial changes are incurred when employing different representations, indicating that providing a comprehensive representation of lesions is a critical factor for effective diagnosis. Also, the color-based features,

i.e., #8~10 and #19~20, achieve better performance on SE on average than the other features, indicating that lesion color is particularly discriminative for diagnosis. Furthermore, the RF classifier shows better performance than the others for color and combined features. This is because the RF incorporates the hierarchical steps of the diagnosis from the dependent lesion properties.

We visualize the diagnosis results of 8 representations for 10 common skin diseases in Figure 4. As shown, single low-level feature are not sufficient to satisfactorily recognizes common skin diseases, while their combinations perform better. When the proposed representations and the combinations are employed, the diagnosis performance is further improved, demonstrating that the proposed representations capture the characteristics of skin lesions.

5.3. Comparison with the CNN and Doctors

The effectiveness of CNNs has been proven in many tasks of computer vision due to their powerful feature representation, especially for classification tasks. Therefore, we also compare the proposed representations with deep features derived from the CaffeNet, VGGNet, GoogleNet and ResNet, the results of which are reported in Table 2. As shown, the deep features derived from the fine-tuned ResNet achieve an ACC of 53.35% and SE of 51.24%,



GT: 5	GT: 46	GT: 31	GT: 1	GT: 15	GT: 15	GT: 92	GT: 56
General D: 5	General D: 67	General D: 68	General D: --	General D: 35	General D: 44	General D: 81	General D: 167
Junior D: 5	Junior D: 46	Junior D: 183	Junior D: 155	Junior D: 35	Junior D: 15	Junior D: 81	Junior D: 56
Expert: 5	Expert: 46	Expert: 55	Expert: 1	Expert: 35	Expert: 15	Expert: 92	Expert: 56
Rank1: 5	Rank1: 46	Rank1: 68	Rank1: 155	Rank1: 35	Rank1: 87	Rank1: 92	Rank1: 56
Rank2: 119	Rank2: 39	Rank2: 178	Rank2: 87	Rank2: 115	Rank2: 124	Rank2: 27	Rank2: 176
Rank3: 49	Rank3: 67	Rank3: 31	Rank3: 1	Rank3: 15	Rank3: 26	Rank3: 115	Rank3: 67

Figure 5. Examples recognized by proposed method and real doctors. Here, ‘GT’ denotes the ground-truth, and each number represents a class of skin disease. The symbol ‘-’ denotes that the doctor cannot provide a diagnosis for this patient. We report the disease predicted according to our method, with the top three probabilities in the last three rows. As is shown, the proposed method achieves comparable performance with the experts, and surpasses the junior doctors for several cases.

Table 2. Comparison to the state-of-the-art deep features derived by different CNNs and the diagnosis of doctors. There are the classification performance of the pre-trained CNN models and those fine-tuned (ft) on SD-198 dataset with the SVM classifier. ‘‘General D’’ is general doctor who does not focus on one specific kind of disease. ‘‘Junior D’’ is junior dermatologist and ‘‘Expert’’ is an expert for diagnosing skin lesions.

Method		ACC	SE
Deep features [33]	CaffeNet	42.31	41.57
	CaffeNet + ft	46.69	45.18
	VGGNet	37.91	37.25
	VGGNet + ft	50.27	48.25
	GoogleNet	35.33	35.21
	GoogleNet + ft	46.48	45.86
	ResNet	48.78	47.62
ResNet + ft	53.35	51.24	
Doctors	General D	49.00	47.50
	Junior D	52.00	53.40
	Expert	83.29	85.00
Ours		56.47	53.15

which is the state-of-the-art performance. In contrast, the proposed representations improve ACC and SE with 3.12% and 1.91%, respectively. The lack of a large-scale well-labeled dataset for this topic limits the application of deep methods, which cannot learn a distinctive embedding based on the existing data. In contrast, the proposed medical representations which are consistent with the observations of dermatologists according to empirical criteria do not rely on such a large-scale dataset. These dermatologist criteria are employed to distinguish diseases by focusing on certain manifestations, based on which the designed features are mapped to a low-rank feature space. Meanwhile, existing deep models rely on the pre-training process using large-scale datasets (ImageNet), which have a large semantic gap regarding the task of skin disease analysis.

In this work, we also report the comparison between the proposed method and three kind of real doctors, *i.e.*, the general, junior and the expert. ‘General doctor’ indicates general doctors with comprehensive knowledge who have

mostly practised in the private clinic and not focus on one kind of disease. ‘Junior dermatologist’ represents the intern dermatologist. ‘Expert’ denotes senior doctors who have considerable experience on the skin disease diagnosis. For each category of doctor we invite two doctors to undertake diagnosis independently followed by a discussion when they are divergent. Table 2 shows the diagnosis results of doctor, and the result of our method is better than the diagnosis of general and junior doctor on SD-198. Note that all the diagnoses of doctors are made by just observing the picture, and they will make better diagnoses by touching the lesion and asking questions. Meanwhile, Figure 5 shows some examples with the diagnosis results from doctors and our method. As is shown, our proposed method shows comparable performance on diagnosing multiple skin diseases compared to real doctors.

6. Conclusions

In this paper we address the problem of clinical skin lesion diagnosis, which is challenging compared to skin cancer recognition on dermatoscopic images. We verify that the criteria employed by clinicians in the diagnosis process can be measured by computers. Accordingly, we design six discriminative and interpretable representations for distinguishing skin lesions by incorporating the accepted dermatological criteria. Experiments on a benchmark dataset demonstrate the proposed representations outperform both the low-level features and the deep features. Furthermore, the final performance on clinical images with 198 categories of skin disease is comparable with dermatologists.

Acknowledgments

This research was supported by NSFC (No. 61620106008, 61572264, 61301238, 61201424), and the Open Project Program of the National Laboratory of Pattern Recognition (NLPR).

References

- [1] N. R. Abbasi, H. M. Shaw, D. S. Rigel, R. J. Friedman, W. H. McCarthy, I. Osman, A. W. Kopf, and D. Polsky. Early diagnosis of cutaneous melanoma: revisiting the ABCD criteria. *The Journal of the American Medical Association*, 292(22), 2004.
- [2] D. G. Altman and J. M. Bland. Diagnostic tests. 1: Sensitivity and specificity. *British Medical Journal*, 308(6943), 1994.
- [3] G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco, and M. Delfino. Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. *Archives of Dermatology*, 134(12), 1998.
- [4] C. Barata, M. A. Figueiredo, M. E. Celebi, and J. S. Marques. Local features applied to dermoscopy images: Bag-of-features versus sparse coding. In *IbPRIA*, 2017.
- [5] G. S. Barsh. What controls variation in human skin color? *PLoS Biology*, 1(1), 2003.
- [6] R. Benavente, M. Vanrell, and R. Baldrich. A data set for fuzzy colour naming. *Color Research & Application*, 31(1), 2006.
- [7] M. Binder, H. Kittler, S. Dreiseitl, H. Ganster, K. Wolff, and H. Pehamberger. Computer-aided epiluminescence microscopy of pigmented skin lesions: the value of clinical data for the classification process. *Melanoma Research*, 10(6), 2000.
- [8] A. P. Dhawan, R. Gordon, and R. M. Rangayyan. Nevoscopy: three-dimensional computed tomography of nevi and melanomas in situ by transillumination. *IEEE Transactions on Medical Imaging*, 3(2), 1984.
- [9] S. Dreiseitl and M. Binder. Do physicians value decision support? a look at the effect of decision support systems on physician opinion. *Artificial Intelligence in Medicine*, 33(1), 2005.
- [10] S. Dreiseitl, L. Ohno-Machado, H. Kittler, S. Vinterbo, H. Billhardt, and M. Binder. A comparison of machine learning methods for the diagnosis of pigmented skin lesions. *Journal of Biomedical Informatics*, 34(1), 2001.
- [11] A. Esteve, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 2017.
- [12] T. Fitzpatrick, J. Bernhard, T. Cropley, et al. The structure of skin lesions and fundamentals of diagnosis. *Dermatology in General Medicine*, 5, 1999.
- [13] A. Gewirtzman, J.-H. Saurat, and R. Braun. An evaluation of dermoscopy fluids and application techniques. *British Journal of Dermatology*, 149(1), 2003.
- [14] J. Glaister, A. Wong, and D. A. Clausi. Segmentation of skin lesions from digital images using joint statistical texture distinctiveness. *IEEE Transactions on Biomedical Engineering*, 61(4), 2014.
- [15] H. M. Gloster and K. Neal. Skin cancer in skin of color. *Journal of the American Academy of Dermatology*, 55(5), 2006.
- [16] A. Green, N. Martin, G. McKenzie, J. Pfitzner, F. Quintarelli, B. Thomas, M. O'Rourke, and N. Knight. Computer image analysis of pigmented skin lesions. *Melanoma Research*, 1(4), 1991.
- [17] T. P. Habif, M. S. Chapman, J. L. Campbell, J. G. Dinulos, and K. A. Zug. *Skin Disease E-Book: Diagnosis and Treatment*. Elsevier Health Sciences, 2011.
- [18] A. Hamilton and R. Brady. Medical professional involvement in smartphone apps in dermatology. *British Journal of Dermatology*, 167(1), 2012.
- [19] H. Handels, T. Roß, J. Kreisusch, H. H. Wolff, and S. J. Poepl. Feature selection for optimized skin tumor recognition using genetic algorithms. *Artificial Intelligence in Medicine*, 16(3), 1999.
- [20] R. J. Hay and L. C. Fuller. The assessment of dermatological needs in resource-poor regions. *International journal of dermatology*, 50(5), 2011.
- [21] R. J. Hay, N. E. Johns, H. C. Williams, I. W. Bolliger, R. P. Dellavalle, D. J. Margolis, R. Marks, L. Naldi, M. A. Weinstock, S. K. Wulf, et al. The global burden of skin disease in 2010: an analysis of the prevalence and impact of skin conditions. *Journal of Investigative Dermatology*, 134(6), 2014.
- [22] M.-L. T. Johnson and J. Roberts. Skin conditions and related need for medical care among persons 1-74 years, United States, 1971-1974. *Vital & Health Statistics*, (212), 1978.
- [23] S. Jowett and T. Ryan. Skin disease and handicap: an analysis of the impact of skin conditions. *Social Science & Medicine*, 20(4), 1985.
- [24] L. S. King. What is disease? *Philosophy of Science*, 21(3), 1954.
- [25] K. Korotkov and R. Garcia. Computerized analysis of pigmented skin lesions: a review. *Artificial Intelligence in Medicine*, 56(2), 2012.
- [26] I. Maglogiannis and C. N. Doukas. Overview of advanced computer vision systems for skin lesions characterization. *IEEE Transactions on Information Technology in Biomedicine*, 2009.
- [27] A. A. Marghoob, L. D. Swindle, C. Z. Moricz, F. A. S. Negron, B. Slue, A. C. Halpern, and A. W. Kopf. Instruments and new technologies for the in vivo diagnosis of melanoma. *Journal of the American Academy of Dermatology*, 49(5), 2003.
- [28] O. Razeghi and G. Qiu. 2309 skin conditions and crowd-sourced high-level knowledge dataset for building a computer aided diagnosis system. In *ISBI*, 2014.
- [29] O. Razeghi, Q. Zhang, and G. Qiu. Interactive skin condition recognition. In *ICME*, 2013.
- [30] J. Spalding. The doctor with an inherited defect of colour vision: effect on clinical skills. *The British Journal of General Practice*, 43(366), 1993.
- [31] W. V. Stoecker and R. H. Moss. Digital imaging in dermatology. *Computerized Medical Imaging and Graphics*, 16(3), 1992.
- [32] W. Stolz, A. Riemann, A. Cognetta, L. Pillet, W. Abmayr, D. Holzel, P. Bilek, F. Nachbar, and M. Landthaler. ABCD rule of dermatoscopy—a new practical method for early recognition of malignant-melanoma. *European Journal of Dermatology*, 4(7), 1994.
- [33] X. Sun, J. Yang, M. Sun, and K. Wang. A benchmark for automatic visual classification of clinical skin disease images. In *ECCV*, 2016.
- [34] S. C. Taylor, F. Cook-Bolden, Z. Rahman, and D. Strachan. Acne vulgaris in skin of color. *Journal of the American Academy of Dermatology*, 2002.
- [35] M. W. Tsang and C. L. Kovarik. Global access to dermatopathology services: physician survey of availability and needs in sub-saharan africa. *Journal of the American Academy of Dermatology*, 63(2), 2010.
- [36] Z. J. Twardowski and B. F. Prowant. Classification of normal and diseased exit sites. *Peritoneal Dialysis International*, 16(Suppl 3), 1996.
- [37] M. Uehara and T. Sawai. A longitudinal study of contact sensitivity in patients with atopic dermatitis. *Archives of Dermatology*, 125(3), 1989.
- [38] R. P. Usatine and M. Riojas. Diagnosis and management of contact dermatitis. *American Family Physician*, 82(3), 2010.
- [39] J. Van De Weijer, C. Schmid, and J. Verbeek. Learning color names from real-world images. In *CVPR*, 2007.
- [40] J. Van De Weijer, C. Schmid, J. Verbeek, and D. Larlus. Learning color names for real-world applications. *IEEE Transactions on Image Processing*, 18(7), 2009.
- [41] J. Zhang, S. Sclaroff, Z. Lin, X. Shen, B. Price, and R. Mech. Minimum barrier salient object detection at 80 fps. In *ICCV*, 2015.