

Accessible Melanoma Detection using Smartphones and Mobile Image Analysis

Thanh-Toan Do, Tuan Hoang, Victor Pomponiu, Yiren Zhou, Zhao Chen, Ngai-Man Cheung, Dawn Koh, Aaron Tan and Suat-Hoon Tan

Abstract—We investigate the design of an entire mobile imaging system for early detection of melanoma. Different from previous work, we focus on smartphone-captured visible light images. Our design addresses two major challenges. First, images acquired using a smartphone under loosely-controlled environmental conditions may be subject to various distortions, and this makes melanoma detection more difficult. Second, processing performed on a smartphone is subject to stringent computation and memory constraints. In our work, we propose a detection system that is optimized to run entirely on the resource-constrained smartphone. Our system intends to localize the skin lesion by combining a lightweight method for skin detection with a hierarchical segmentation approach using two fast segmentation methods. Moreover, we study an extensive set of image features and propose new numerical features to characterize a skin lesion. Furthermore, we propose an improved feature selection algorithm to determine a small set of discriminative features used by the final lightweight system. In addition, we study the human-computer interface (HCI) design to understand the usability and acceptance issues of the proposed system. Our extensive evaluation on an image dataset provided by National Skin Center - Singapore (117 benign nevi and 67 malignant melanoma) confirms the effectiveness of the proposed system for melanoma detection: 89.09% sensitivity at specificity $\geq 90\%$.

Index Terms—Multimedia-based healthcare, malignant melanoma (MM), mobile image analysis, feature selection, Human-Computer Interface.

I. INTRODUCTION

MOBILE devices, including smartphones, are being used by billions of people all around the world. This creates the opportunity to design a wide variety of mobile image applications, e.g., mobile image search [1]–[3], landmark recognition [4], [5], mobile video type classification [6] and 3-D scene video [7]. Among many imaging applications, healthcare applications have drawn a lot of attentions recently. Several methods [8]–[11] have been proposed to support efficient and timely image-related diagnosis. Apart from normal imaging healthcare applications, mobile imaging healthcare applications have the advantages of being practical, low-cost and accessible [12]–[14].

The work focuses on accessible detection of malignant melanoma (MM) using mobile image analysis. MM is a type of skin cancer arising from the pigment cells of the epidermis. There are three main types of skin cancers: MM, basal cell carcinoma and squamous cell carcinomas. Among them, MM

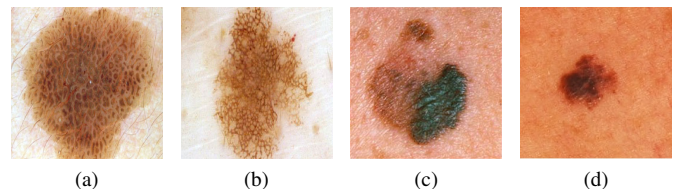


Fig. 1: Comparison of dermoscopic images (a, b) and visible light images (c, d) of skin lesions. Dermoscopic images are taken with the aid of liquid medium or non-polarized light source and magnifiers, and they include features under the skin surface (e.g., pigment network, aggregated globules). On the contrary, visible light images (e.g., taken with smartphones) do not include these features. This work focuses on the analysis of visible light images. Source: www.dermoscopy.org, NSC.

is considered most hazardous. According to an annual [15], the American Cancer Society projected 87,110 new cases of melanoma in the United States by the end of 2017, with almost 9,730 estimated deaths. MM may be treated successfully, yet the curability depends on its early detection and removal when the tumor is still relatively small and thin. Therefore, there is a pressing need for tools that can assist early and accurate diagnosis.

The current practice of initial melanoma diagnosis is clinical and subjective [16], relying mainly on the use of naked eye examination. Therefore, the diagnostic accuracy is highly dependent on the trained expertise of dermatologists, which is estimated to be about 85% [17].

In order to improve the diagnostic identification of MM, dermatologists may use other visual aids such as dermatoscopy [17]. In clinical evaluation, dermatologists typically apply the ABCDE signs [18] (which stands for Asymmetry of the lesion, Border irregularity, Color variation, Diameter, and Evolving), the 7-point checklist [19], and the Menzies method [20], followed by biopsy confirmation of diagnosis. While diagnosis by dermatologists is accurate, a clinic visit may be less easily accessible and may require the primary physician to make the initial referral. There is a need for the public to be educated and equipped with a more accessible method of self-assessment for early diagnosis of melanoma.

Nowadays our industry faces the junction of two rapidly developing markets: healthcare and emerging mobile computing. The ever-increasing availability of mobile devices equipped with multi-core CPUs and high resolution image sensors have the potential to empower people to become more proactive and engaged in their own healthcare processes [21].

T.-T. Do, T. Hoang, V. Pomponiu, Y. Zhou, Z. Chen, N.-M. Cheung and D. Koh are with the Singapore University of Technology and Design (SUTD). A. Tan and S.-H. Tan are with National Skin Center (NSC), Singapore
Corresponding author: N.-M. Cheung, SUTD. Email: ngaiman_cheung@sutd.edu.sg

Our contribution is the design of a complete mobile imaging system to detect melanoma. Our system uses an ordinary smartphone as the platform. Thus, the proposed system is remarkably accessible. Our system employs state-of-the-art image analysis algorithms to enable automatic assessment of the malignancy of the skin lesion. Our goal is that the general public can use our proposed mobile health (mHealth) system to perform preliminary assessment frequently and detect any anomalous skin lesion in their early stage.

As will be further discussed, the proposed system has four major components. The first component is a fast and lightweight segmentation algorithm for skin lesion localization. The second component incorporates new computational features to improve detection accuracy for smartphone-captured skin lesion images. The third component applies new feature selection tools to select good features used for the classification. The fourth component includes a classifier array and an algorithm to fuse classification results. An iterative design approach is used to assess and improve the performance and utility of the proposed system. We extensively study the system in pre-clinical settings, based on a large number of MM and benign nevi images.

Note that using visible light images captured from smartphones for automatic melanoma detection is quite new. Most previous works focused on dermoscopic images that are captured in the well-controlled clinical environments with specialized equipments [22]–[24] (including [25], which uses a smartphone for dermoscopic image analysis). Dermoscopic images include features below the skin surface, which cannot be captured with normal cameras equipped in smartphones (Fig. 1). In addition, smartphone-captured images may be subject to various types of distortion (illumination variation, motion blur, defocus aberration). Therefore, melanoma detection using smartphone-captured visible light images poses some unique challenges, and the problem is not well-understood [26], [27]. Furthermore, in order to perform the entire image analysis on the smartphone, we need to design efficient algorithms and system feasible for the strict computation, memory and power constraints of a smartphone. To address all these challenges, our works make the following novel contributions [28]:

- We propose a light-weight skin lesion localization algorithm suitable for the resource-constrained smartphone. Our localization algorithm comprises skin / non-skin detection, hierarchical segmentation, and combination of Otsu's method and Minimum Spanning Tree (MST) method.
- We use novel color and border features to quantify the color variation and the border irregularity of skin lesions. We evaluate 116 computational features to quantify the color, border, asymmetry and texture of a skin lesion, including our proposed features that are suitable for visible light images captured under loosely-controlled lighting conditions.
- We investigate feature selection to identify a small set of the most discriminative features to be used in the smartphone. Using a small set of discriminative features not only reduces the storage and computation overhead but also improves the classification performance, as low

dimensional feature vector is more robust to over-fitting. We focus on the framework using normalized mutual information and propose an improvement that takes into account the feature coordinates.

- We propose several methods to fuse the classification results of individual category classifiers.
- We evaluate our system using a dataset from National Skin Center (NSC) of Singapore.
- We study the Human Computer Interface (HCI) aspect of the proposed system.

The remaining sections of the paper are structured as follows. Section II reviews melanoma analysis methods. Section III presents the details of our proposed system and algorithms. Section IV presents the evaluation and experiments. Finally, Sections V and VI discuss the limitation of the system and conclude the paper by highlighting several future research directions. A preliminary version of this work has been reported [28], [29]. This paper discusses substantial extension to our previous work: (i) We improve and quantitatively evaluate the proposed segmentation method (Section IV-B). (ii) We employ and evaluate the Local Binary Pattern descriptor [30] for the skin lesion analysis (Section IV-D). (iii) We propose new methods for fusing the results of individual category classifiers (Section III-D). The experimental results show that the proposed fusion methods boost the classification accuracy. (iv) We compare the proposed system to the recent melanoma detection methods [31], [32] (Section IV-E). (v) We also study the human-computer interface (HCI) aspect of the proposed system (Section IV-F).

II. RELATED WORKS

Depending on the mechanism used to evaluate the skin lesion, melanoma diagnosis schemes can be classified into the following classes: manual methods [20], which require the visual inspection by an experienced dermatologist, and automated (computed-aided) schemes [22], [25], [33], that perform the assessment without human intervention. A different class, called hybrid schemes, can be identified when dermatologists jointly combine the computer-based result, context knowledge (e.g., skin type, age, gender) and his experience during the final decision [34].

In general, an automatic melanoma analysis system can be constructed in four main phases. The first phase is the image acquisition which can be performed through different devices such as dermatoscope, spectroscopy, standard digital camera or camera phone. The images acquired by these devices exhibit peculiar features and different qualities, which can significantly change the outcome of the analysis process. The second phase involves skin detection, by removing artifacts (e.g., ruler, hair), and lesion border localization. The third phase computes a compact set of discriminative features. Finally, the fourth phase classifies the lesions based on the extracted features.

There is a plethora of computer-aided systems for segmentation and classification of skin lesions [22], [23], [33], [35], [36]. Most of these works investigated for lesion segmentation of a dermoscopic image by using classic image segmentation

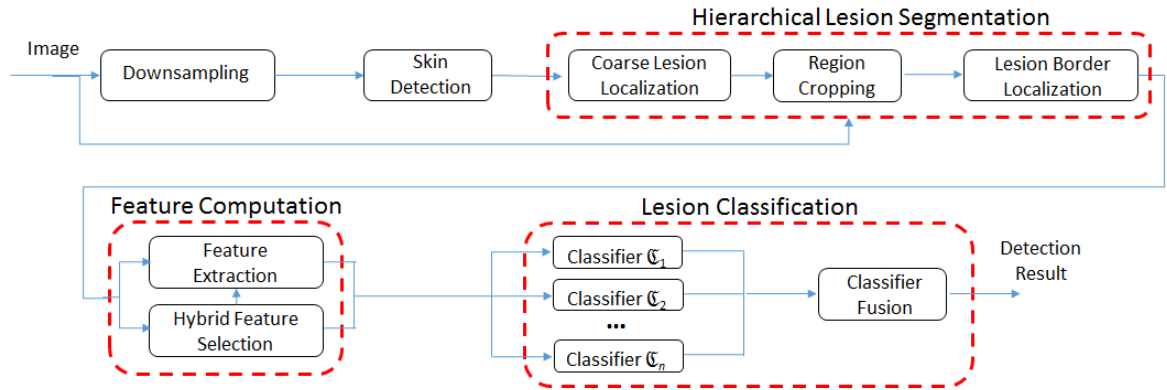


Fig. 2: The flowchart of the proposed method.

methods such as histogram thresholding, adaptive thresholding, difference of Gaussian filter, morphological thresholding, wavelet transform, active contours, adaptive snake and random walker algorithm. However, it is worth noting that these works focused on dermoscopic images which are acquired under controlled clinical conditions by employing a liquid medium (or a non-polarized light source) and magnifiers. This type of images includes features below the skin surface which cannot be captured with standard cameras. It is also worth noting that many of these previous works focused on only a certain aspect (e.g., lesion border localization), and they did not provide a complete solution that integrates all the steps. In addition, they focused on processing on powerful workstations / servers, where computation and memory are abundant. In the cases of redundant memory and computation capability, recent dermoscopic images-based melanoma detection leverage the power of deep neuron network (DNN) as this could help to achieve very competitive performances [37], [38].

Recently, several mobile connected dermatoscopic devices have been developed, such as DermLite (3Gen Inc, USA) and HandyScope (FotoFinder Systems, Germany). However, the cost to acquire such an additional device is expensive and they are not accessible to everyone. Furthermore, trained personnel are required to operate dermatoscopic devices.

There are only few systems working on mobile platforms: Lubax (Lubax Inc, CA, USA), [25], [27], [39], [40]. However, many methods merely use the mobile device for capturing, storing and transmission of the skin lesion images to a remote server without performing any computation locally on the mobile device. For example, [39], [40] use a mobile dermatoscope attached to the mobile device to capture dermatoscopic images and send the images to the server for computer assessment.

A few isolated works perform the analysis of smartphone-captured visible-light images directly on the mobile devices. In [26], a mobile-system working for images taken from mobile cameras is presented. In particular, they presented a preliminary system: to detect a lesion, they used a very basic thresholding method; to describe the lesion, only standard color feature (such as mean and variance of the color channels, the difference of color through vertical axis) and border features (convexity, compactness) are extracted. In [27], the authors also focused on images taken from mobile cameras.

The lesion detection and feature extraction are performed on mobile while the classification can be performed on mobile or cloud. However, the authors put more emphasis on the system integration, without mentioning the details of the features used for diagnosis. Recently, the works [31], [41] propose complete systems that segment, extract visual features and classify lesions. In [31], in addition to the automatic extracted color and texture features, additional human annotated information, including lesion's locations, lesion size, number of lesions, etc. is also utilized to differentiate melanoma from nevocellular naevi. The work [41] only uses the automatically extracted features, which are based on the ABCD rule, i.e., Asymmetry, Border, Color, and Diameter features, for the classification process. Different from these works, which only use general features such as mean, variance of different color channels, convexity, solidity, compactness of shape etc., in our work, we propose novel and robust features specifically for characterization of lesions. We also propose a novel feature selection method to select a small but very discriminative set of features which not only helps boost the classification accuracy but also reduce the computation and memory costs. Additionally, even though the works [31], [41] propose complete systems, the efficiency of these systems running on resource-constrained smartphones has not been reported. On the other hand, our method is very efficient on a smartphone as discussed in Section IV-F.

Recently, there are several DNN-based systems that have been proposed for non-dermatoscopic image analysis: lesion segmentation [42] and melanoma detection [43], [44]. However, due to the high computation and memory cost of DNN, it is very challenging for such systems to be used on resource-constrained smartphone platforms.

III. THE METHOD FOR EARLY MELANOMA DETECTION

Fig. 2 depicts our proposed scheme. It is challenging to achieve accurate segmentation of skin lesions from smartphone-captured images under loosely controlled lighting and focal conditions. Instead of using sophisticated segmentation algorithms, which can be computationally expensive, we propose to localize the skin lesion with a combination of skin detection and a fast hierarchical segmentation. More precisely, we first downsample the skin image, and based on

the downsampled version our system generates a coarse model of the lesion by combining two lightweight segmentation algorithms. Then, to outline the lesion contour, we employ a fine segmentation by using as input the coarse segmentation result. From the final segmented region, we extract four feature categories which accurately characterize the lesion color, shape, border and texture. To classify the skin lesion, a classifier is built for each feature category and then the final results is obtained by fusing their results.

A. Lesion segmentation

Our segmentation process consists of two main steps. At the first step, a mask of skin regions is generated using the skin detection method. The purpose of the skin detection module is to discard pixels from non-skin regions to simplify the image for the subsequent processing step. At the second step, we extract the lesion by using a hierarchical segmentation method.

1) *Skin Detection*: The reason for applying a skin detection procedure first is to filter the image from unwanted artifacts, so an exact classification of skin/non-skin regions are not needed as long as we extract the foreground and keep the whole lesion region within. Here we use an approach based on skin color model to detect skin pixels [45]. We choose this particular skin model since it is more discriminative, providing 32 skin and non-skin color maps of size $64 \times 64 \times 64$ for each skin color. We use the original RGB color image, without any preprocessing, as input to the skin detection model.

In order to build the skin detection model, we followed the steps given in [45]: we first collected a set of skin/non-skin images to construct our skin detection dataset. Skin images are selected with different skin colors and various lighting conditions for model generalization. The skin color distribution is estimated by a Gaussian mixture model. Since the skin mole we want to localize may not have a similar color as the surrounding skin, we use a filling method for all the holes inside the skin region.

2) *Hierarchical Lesion Segmentation*: Since our objective is to develop a mobile-based diagnosis system, we need a lightweight segmentation method that can achieve high precision under severe computational constraints. Therefore, for the segmentation engine we employ several segmentation methods with low computational requirements, followed by the use of a criterion to merge the results. It worth noting that the skin lesion images are converted to the grayscale space for the hierarchical segmentation.

a) *Coarse Lesion Localization*: After getting the skin region area we downsample the image and perform two segmentation methods and use rules to combine the results of both methods. Here we select Otsu's method [46] and Minimum Spanning Tree (MST) method [47] to get the initial segmentation results. Fig. 3 shows the flowchart of the coarse lesion localization procedure.

Otsu's method [46] is a simple and fast thresholding method that automatically calculates a threshold value from image histogram. The threshold value is then used to classify image pixels based on their intensities. However, Otsu thresholding may not detect clear edges on the image, for example, the lesion boundaries.

On the other hand, the Minimum Spanning Tree (MST) method [47], a fast region-based graph-cut method, is sensitive to clear edges. However, MST is not robust to detect smooth changes of color intensity. In our method, the parameters of the MST are optimized such that we could get enough candidate region of interest (ROIs) while avoiding over-segmentation near the skin mole region. As segmentation-based MST algorithm that can run at nearly linear time complexity in the number of pixels, we achieve a low time complexity after running two different segmentation methods, i.e., Otsu's and MST methods.

To filter the segmentation results, we firstly remove all candidate ROIs that are connected to the boundary of skin image. In addition, we assume that the skin mole is located in a region (called the valid region) near the center of the image. This hypothesis is adopted since most of the users focus their camera phone on the object of interest (i.e., the skin mole) when capturing a picture. As a consequence, all the candidate ROIs that have the centroid coordinates outside the valid region are discarded. Furthermore, we use a constraint to further discard the noisy ROIs. Specifically, for each segmentation method, we identify the ROI by:

$$\underset{i=1, \dots, n_R}{\operatorname{argmax}} \left\{ A_i \cdot \left[1 - 2 \cdot \sqrt{\left(\mathfrak{C}_i^x - \frac{W}{2} \right)^2 + \left(\mathfrak{C}_i^y - \frac{H}{2} \right)^2} \right]^4 \right\} \quad (1)$$

where, for the i^{th} ROI, A_i denotes its area, and \mathfrak{C}_i^x and \mathfrak{C}_i^y are centroid coordinates. W and H are the width and the height of the downsampled image, respectively. n_R represents the total number of ROIs that are located in the valid region. The basic idea is to give central ROIs high weights while penalizing ROIs near to boundary. When both \mathfrak{C}_i^x and \mathfrak{C}_i^y are close to the image center, then the value $\left(1 - 2 \cdot \sqrt{\left(\mathfrak{C}_i^x - W/2 \right)^2 + \left(\mathfrak{C}_i^y - H/2 \right)^2} \right)^4$ is close to 1. On the other hand, if the ROI is far from the image center, that value is small. The quantity is raised to the fourth power to impose penalty for boundary ROIs.

By merging the two filtered segmentation results, we expect to get a good segmentation on lesion with either distinct border or diffuse border. Based on the rules to perform a fusion of different segmentation given in [23], we take the union of the two results and then find the largest connected region in the union result. The fused segmentation result is post processed by applying a majority filter to remove the noise.

b) *Border Localization*: Given the coarse segmentation result, in order to correctly localize the border of the lesion, we first crop the corresponding ROI from the original high-resolution image. As downsampling performed in the "Coarse Lesion Localization" step may smear the mole boundary, this mapping is not exact and generates some uncertainty related to mole boundary. Thus, another fine-grained segmentation operation is performed to improve the lesion border localization. The segmentation algorithm used for this stage is similar to the one presented in the previous section (i.e., applying Otsu and MST, filtering near boundary segments, and fusing filtered segments), except that we adapt the segmentation parameters to the cropped image characteristics, i.e., a skin

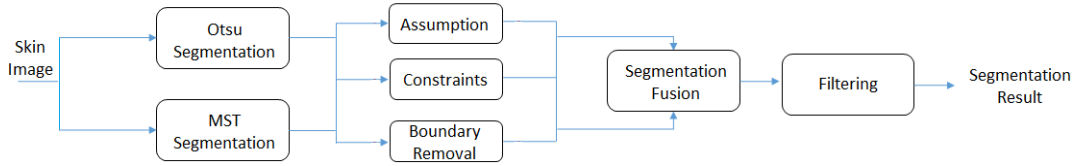


Fig. 3: The block diagram of the coarse lesion localization.

image containing a mole which occupies a large part of the image.

B. Feature Descriptors for Describing Lesion

Given the lesion image segmented from section III-A, we compute features belonging to four categories (color, border, asymmetry and texture) to describe the lesion. The summary of features is given in Table I. Detail of features are presented in the follows.

1) *Lesion Color Feature (LCF) (54 Features)*: Given a skin lesion, we calculate the color features widely used in the literature such as mean, variance of the pixel values of several color channels. The used color channels were *red*, *green*, *blue* from the RGB image; *gray scale* from the gray image, and *hue* and *value* from the HSV image.

To capture more color variation from the skin lesion, we use the information from the histogram of pixel values [22], [23]. A histogram, with fixed number of bins (i.e. 16 bins), of the pixel values in the lesion is computed and the number of non-zero bins is used as the discriminative feature. The features generated from the corresponding color channels are called *num_red*, *num_green*, *num_blue*, *num_gray*, *num_hue* and *num_value*. As melanoma samples usually have higher color variation than benign samples [18], we expect that the number of non-zero histogram bins of melanoma samples is higher than that of the benign samples.

Novel feature to quantify color variation: Generally, there are different color distribution patterns over a MM lesion, whereas the normal pigmented nevi exhibit a more uniform color representation. Therefore, another measurement is needed to determine if there is a color variation all over the lesion or the color varies uniformly from the center to the border. Inspired by the clinical research [24], we use a feature called Color Triangle (CT) [28]. Specifically, given a lesion region (represented by a binary image), we first compute the center of mass of the region. Let N be the number of boundary pixels of the region. Starting from the left most boundary pixel, we divide the region boundary into PA parts in which the number of boundary pixels of each part is equal (i.e. each part has N/PA boundary pixels). Each triangular part is now represented by two segment lines (from the lesion region center to the boundary). We further equally divide each segment line into SP parts. By connecting corresponding parts on two segment lines, we get SP subparts for each triangular region.

After that, each triangular part is described by a SP -component vector, where each component is the mean pixel

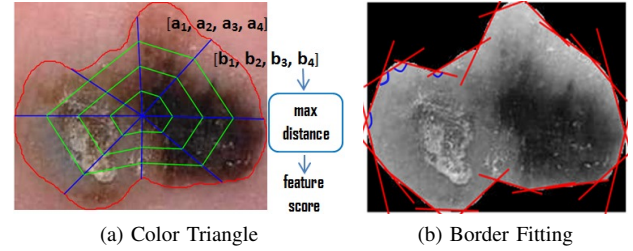


Fig. 4: Novel features to quantify color variation (a): Color Triangle (CT): $SP = 4$ and $PA = 8$ and border irregularity: $nt = 20$ (b) of a skin lesion [28]

values of the subpart [28]. Finally, the maximum Euclidean distance between the vectors is used to quantify the color variation of the skin lesion. As the description, the Color Triangle features are designed to assess the color spreading from the lesion center to the lesion boundary. As melanoma mole has non-uniform color spreading [18], we expect that when the CT feature is large, the chance of melanoma is high.

In our experiments, the values of PA are empirically chosen as 4, 8, 12 and 16. For each value of PA , values of SP are chosen as 2, 4 and 8. The Color Triangle features are computed for *gray scale*, *red* and *hue* channels of the lesion. Fig. 4(a) illustrates for the proposed Color Triangle feature in which $PA = 8$ and $SP = 4$. Totally, we extract 54 color features to describe the color variation. This original feature set is subject to feature selection subsequently to identify a small subset of discriminative features for mobile implementation.

Note that we have included mean/variance of the color channel as the candidates in the original feature set. While we expect our new color feature (Color Triangle) to be more discriminative, we have also included mean/variance in the original feature set (before feature selection) for these reasons: 1) mean/variance are conventional features and we have included them as baseline for comparison (during feature selection); 2) mean/variance do not have any parameter to decide, potentially they could be rather general and robust. On the other hand, for histogram or Color Triangle, we need to decide the number of bins / partitions. Nevertheless, as will be discussed, our feature selection results suggest that the generality of mean/variance cannot compensate for their inadequacy in discriminativeness in this application, and mean/variance are not selected in our final feature set.

2) *Lesion Border Feature (LBF) (16 Features)*: To describe the irregularity of the border, we compute shape features such as compactness, solidity, convexity, and variance of distances from border points to centroid of lesion [24]. Border irregu-

larity is useful for melanoma detection: mole with irregular border has a higher probability of being melanoma [18].

Novel feature to quantify border irregularity: We also propose the Border Fitting feature to quantify the border irregularity [28]. The main idea is to approximate the lesion contour by multiple straight lines and then to use the angles between these lines to quantify the irregularity. Regular borders tend to have smooth and consistent changes between consecutive contour fitting lines, compared with the irregular ones.

We use a linear regression model to determine the contour fitting lines. We partition the contour (border) points of the lesion into nt segments and $\{x_k, y_k\}$ are the corresponding coordinates for the i -th segment. In the linear regression, the predicted coordinate \hat{y}_k can be expressed as $\hat{y}_k = a_0^i + a_1^i \cdot x_k$, where a_0^i and a_1^i are the slope and intercept of the regression line for the i -th contour segment.

To quantify the border irregularity, we compute the angles between consecutive regression lines (i.e., regression lines for the i -th contour segment and the $(i + 1)$ -th contour segment). Then, we compute the average and the variance of the angles, and use these statistics as the border irregularity features. In our experiments, number of lines nt are chosen as 8, 12, 16, 20, 24 and 28. Fig. 4(b) illustrates the Border Fitting feature in which $nt = 20$. Totally, we extract 16 features to describe the border irregularity.

Note that we have included other conventional border features in the original feature set, e.g., convexity, as baseline for comparison (during feature selection), but we expect that they are less discriminative for border irregularity. Our proposed new feature Border Fitting can capture border irregularity well as we directly fit line segments on the border of the mole, and we compute mean / variance of angles between consecutive line segments. In particular, Border Fitting can capture notches on the mole border and is useful for melanoma detection [18].

3) *Lesion Asymmetry Feature (LAF) (1 Feature):* The lesion asymmetry can also reveal valuable information for the lesion categorization. According to the analysis in [24], melanoma samples tend to be asymmetric, while benign samples tend to be symmetric. To compute the lesion asymmetry, we use a method similar to the one introduced in [24]. The major and minor axes of lesion region, i.e., the first and second principal components, are determined. The lesion is rotated such that the principal axes are coincided with the image axes x and y . The object was hypothetically folded to the x -axis and the area difference i.e., A_x , between the two parts was taken as the amount of asymmetry corresponding to the x -axis. We followed the same procedure, for the y -axis, to obtain A_y . The asymmetric feature is computed as $Asym = (A_x + A_y)/A$, where A is the lesion area.

4) *Lesion Texture Feature (LTF) (45 Features):* To quantify the texture of the skin lesion, we investigated several feature descriptors such as:

- those derived from the gray level co-occurrence matrix (GLCM) and
- those based on the local binary patterns (LBP).

The GLCM of the entire lesion characterizes the texture by calculating how often pairs of the pixel with specific brightness values and orientation occur in an image. GLCM-based texture

Feature Category	Used features
LCF	Mean, variance of different color spaces (RGB, HSV and grayscale).
	Number of non-zero bins of the histograms of different color spaces (RGB, HSV and grayscale).
	Color Triangle feature.
LBF	Shape features such as compactness, solidity, convexity
	Variance of distances from lesion border points to the lesion centroid.
LAF	Border Fitting feature.
LTF	Shape asymmetry.
	Edge density.
	Features computed from the GLCM [48] like energy, correlation, contrast, entropy.
	Rotation invariant sign and magnitude of LBP [50] of different color spaces (RGB and grayscale).

TABLE I: Computed features according to their relevant categories and their corresponding descriptions.

description is one of the most well-known and widely used methods in the literature [22], [48].

In this work, GLCM is constructed by considering each two adjacent pixels in the horizontal direction. The features extracted from GLCM used to describe the lesion are contrast, energy, correlation, and homogeneity. To achieve a reasonable estimation of the features, the GLCM should be a dense matrix. Hence, before GLCM calculation, the pixel values are quantized to 32 and 64 levels. It means that we computed 8 texture features from two quantized GLCMs.

To capture edge map (structure) of the lesion, we employed the Canny edge detector method. The number of edge pixels is counted and normalized by total lesion area and the resulted number is used as an edge density feature.

Another widely used texture descriptor that we employed for skin lesion analysis is Local Binary Pattern (LBP) [30]. LBP combines shape and statistical information by a histogram of LBP codes which resemble microstructures in the image. The LBP is a scale invariant measure that describes the local structure in a 3×3 pixel block [49]. The LBP was further adapted to accommodate arbitrary block sizes [30]. Two main parameters of LBP are P and R , i.e., the number of pixel neighbors P on a circle of radius R of a given center pixel. To generate rotation invariant LBP_S, $P-1$ bitwise shift operations of the circle are performed, and the smallest value is selected. In this application, we adopt the LBP framework introduced in [30], since it has a complete mathematical formulation of the LBP operator and it has been extensively tested, offering the best performance. Following [30], we extract the sign LBP (LBP_S) from the grayscale channel of the lesion image. The resulted LBP_S is a 36-dimensional vector.

C. Feature Selection

Given the feature set \mathcal{F} and the class label L , we perform feature selection offline to identify a set $\mathcal{S} \subset \mathcal{F}$ ($|\mathcal{S}| \ll |\mathcal{F}|$) such that the relevance between L and \mathcal{S} is maximized. The relevance is usually characterized in terms of Mutual Information (MI) [51]–[53]. Considering all possible feature subsets requires an exhaustive search which is not practical for a large feature set.

In our method, we use the well-known feature selection procedure called Normalized Mutual Information Feature Selection (NMIFS) [53]. Mutual information is widely employed for the feature selection problem to capture the relevance between variables. In NMIFS, features are selected one by one from \mathcal{F} iteratively. Initially, the feature that maximizes relevance with target class L is selected as the first feature. Given the set of selected feature $\mathcal{S} = \{f_s\}, s = 1, \dots, |\mathcal{S}|$, the next feature $f_i \in \mathcal{F} \setminus \mathcal{S}$ is chosen such that it maximizes the relevance of f_i with the target class L and minimizes the redundancy between it and the selected features in \mathcal{S} . In other words, f_i is selected by the following condition:

$$\operatorname{argmax}_i \left\{ MI(L, f_i) - \frac{1}{|\mathcal{S}|} \sum_{f_s \in \mathcal{S}} NMI(f_i, f_s) \right\} \quad (2)$$

where $MI(X, Y)$ is the mutual information, which measures the relevance between two random variables X and Y and is defined as

$$MI(X, Y) = \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (3)$$

while NMI is the normalized mutual information and is defined as

$$NMI(X, Y) = \frac{MI(X, Y)}{\min\{H(X), H(Y)\}} \quad (4)$$

where H is the entropy. From information theory, it is known that $MI(X, Y) \geq 0$; if X or Y is binary variable then $MI(X, Y) \leq 1$; and we always have $0 \leq NMI(X, Y) \leq 1$.

Novel feature selection criterion [28]: From (3) we observe that the mutual information is a metric that relies on the probability functions, and it is independent of the coordinate of the features. On the other hand, the coordinate of the features may help in the classification context. For example, in a binary classification problem, suppose that the number of samples in each class are equal and there are two features f_i and f_j which perfectly separate the two classes. The feature that has a larger margin between the two classes has a better generalization error. However, by using mutual information, it can be seen from (3) that these features have the same MI value (Fig. 5). Furthermore, by using the feature coordinates, we can access the local neighborhood structure of data. This also helps the classification [54]. Therefore, we take into consideration the feature coordinates when selecting the feature.

A well-known criterion considering the coordinate of features is Fisher criterion. Nevertheless, there are several issues with the Fisher criterion [55]. In particular, if the data in each class does not follow a Gaussian distribution, or if the mean values of the classes are approximately equal, Fisher criterion fails.

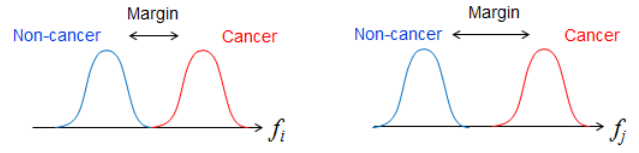


Fig. 5: Two features f_i and f_j that can perfectly separate the two classes. Feature f_j has a larger margin between the two classes and has a better generalization error. However, in conventional feature selection, these features have the same MI value following (3), as feature coordinates are not considered.

Inspired by the method “Average Neighborhood Margin (ANM) maximization” [54] method proposed for face recognition problem, we adapt that method to the feature selection problem and define the quality of feature f as

$$Q(f) = \sum_{i=1}^n \left\| \left(\sum_{t \in n_i^e} \frac{\|f(i) - f(t)\|_1}{n_i^e} - \sum_{j \in n_i^o} \frac{\|f(i) - f(j)\|_1}{n_i^o} \right) \right\|_1 \quad (5)$$

where, for each sample i , n_i^o is the set of the most similar samples which are in the same class with i and n_i^e is the set of the most similar samples which are not in the same class with i [28]. In (5) a feature has good discriminative power if we can use it to separate each sample from the samples belonging other classes whilst it is close to samples belonging to the same class. Since (5) makes use of local information and does not make any assumptions on the distributions of samples, it can overcome the drawbacks of the Fisher test.

In order to combine the advantages of both (5) (which works on the local neighborhood structure of data) and MI (which works on the distribution of data), we propose the new criterion which combines both (5) and MI in a single criterion as follows:

$$U(f_i) = \alpha \cdot Q(f_i) + (1 - \alpha) \cdot \left\{ MI(L, f_i) - \frac{1}{|\mathcal{S}|} \sum_{f_s \in \mathcal{S}} NMI(f_i, f_s) \right\} \quad (6)$$

where $\alpha \in [0, 1]$ is a weight factor that controls the influence of (5) and MI in the proposed hybrid criterion [28].

D. Classifier

The offline feature selection process is performed individually for each feature category. A compact set of features is identified for each category (color, border, asymmetry, texture: GLCM features and edge density feature). For a given lesion image, our system computes these selected features for each category and passes them into SVM [56] classifiers. We train four different SVM classifiers, each corresponding to a single feature category. In this application, we apply a higher penalty for each misclassification of MM samples than each of benign samples (i.e., the penalty is 1.5 and 1 for MM and benign, respectively) as we want to achieve high sensitivity while maintaining a reasonable specificity.

For the LBP features, we use a different classifier. The k nearest neighbor classifier (k NN) has been applied for

classification of a majority of LBP descriptors [57] since it is a good candidate when working in a distance representation of objects. Furthermore, since LBP produces a high dimensional feature vector, we will require a large number of samples in order to project the LBP features to higher space where we can use SVM. As a consequence, we adopt k NN classifier. In general, several metrics have been employed to compute the distance between the LBP histograms such as chi-square distance, L1 distance, and cosine distance. In this study, we adopt the cosine distance metric for the k NN classifier since it is more robust to outliers and it has been widely used in many previous works [57].

Additionally, it is very helpful to fuse the results from individual classifiers as this combines valuable information gained in the training phase of each feature category. Specifically, we propose and experiment these fusion methods:

- 1) *Non-weighted sum*: We sum the hard classification results of each classifier (1: cancer or 0: non-cancer). By using the sum rule, we base on the assumption that each feature category contributes equally during the diagnosis decision. Consequently, a skin lesion is judged as cancerous if the sum of the four hard classification values is ≥ 1 [28].
- 2) *Weighted sum*: Potentially, individual feature categories have different discriminative power. Therefore, we fuse their classification results by summing the weighted hard classification values. Particularly, we compute the weights in two different ways: (2a) *sensitivity@50% specificity* and (2b) *the Area Under Receive Operating Curve (AUC)*. We compute these quantities for individual classifiers during the validation phase, and apply these quantities as weights during the testing phase. For this method, a threshold on the sum of weighted hard classification values is required to decide if a sample is cancerous. We select the threshold that maximizes the accuracy in the validation set.
- 3) *Hierarchical SVM*: We concatenate the soft classification values of individual classifiers, resulting a new feature vector. The new feature vector is subjected to another SVM for classification. We call this fusion method as hierarchical SVM. This fusion method can uncover the underlying non-linear combination of features in the diagnosis.

Note that we have two techniques to extract skin lesion texture: GLCM plus edge density, and LBP. Thus, we experiment these fusion methods for two sets of feature categories: (i) color, border, asymmetry, GLCM plus edge density and (ii) color, border, asymmetry, LBP.

E. Human Computer Interface (HCI) Design

We also study the HCI aspects of the proposed system. The proposed system can be regarded as *self-diagnosis* applications [58]. Self-diagnosis applications refer to a type of accessible healthcare systems that can help users detect illness at the early stages. Users of self-diagnosis applications are non-patients: people who do not have the explicit awareness of their potential diseases or health problems. For these self-diagnosis

applications, it is not clear how to incentivize usage, how to best present the results, and how to sustain habit formation for periodic long-term use. To understand these issues, we conduct an exploratory case study to investigate the special HCI design challenges of such systems. In particular, we conduct a semi-structured interview with 16 healthy young people and 13 healthy older adults to investigate the specific HCI challenges of our proposed application.

IV. EVALUATION AND RESULTS

A. Dataset and evaluation protocol

To evaluate the proposed scheme, we use the dataset provided by the National Skin Center (NSC) of Singapore. There are totally 184 color images (**SET1**) consisting of two classes: 117 images of benign nevus and 67 images of MM, acquired by cameras under different resolutions, sizes, and conditions. Due to various acquisition conditions (such as lighting and focus) and presences of other anatomical features (e.g. eye, eyebrow, nail, etc.) near the skin lesions, many of these images are challenging for the segmentation and classification.

In regards to the ground truth for both segmentation and classification tasks, we obtained the diagnosis of the melanoma cases which were determined by histopathological examination or clinical agreement by several expert dermatologists from NSC. Furthermore, an expert was engaged in the annotation process to help achieve the reliable ground truth ROIs for the skin lesions.

A smaller dataset called **SET2**, which is a subset of SET1, i.e. 52 benign images and 29 MM images, is used as the working data for the feature selection phase.

Additionally, we partition SET1 into two subsets. The first subset, called **SET3**, consists of 30 images, i.e., 18 benign nevi and 12 MM. This subset is used as working data to determine the weight values for the fusion method 2 proposed in section III-D. The second subset, called **TEST SET**, consists of 154 images, i.e., 99 benign nevi and 55 MM. We perform 10-folds cross validation on this TEST SET to obtain the final classification results. We evaluate the classification performance in terms of sensitivity (i.e., $\text{Sens} = \text{TP}/(\text{TP} + \text{FN})$), specificity (i.e., $\text{Spec} = \text{FN}/(\text{FN} + \text{FP})$) and balanced accuracy (i.e., $\text{Acc} = (\text{Sens} + \text{Spec})/2$).

B. Segmentation Results

The hierarchical segmentation process discussed in Section III is applied on the SET1 to extract lesion ROIs.

To measure the segmentation results, we used the true detection rate (TDR), which quantifies the rate of pixels correctly classified as lesion. The TDR is computed as follows:

$$\text{TDR}(\text{GT}, \text{SEG}) = 100 \cdot \frac{\#(\text{GT} \cap \text{SEG})}{\#(\text{GT})} \quad (7)$$

where SEG denotes the segmentation result and GT denotes the ground truth segmentation.

Table II presents segmentation results of Otsu [46], MST [47], and the proposed method. It clearly shows that the proposed hierarchical segmentation significantly improves the segmentation results over Otsu and MST methods.

Method	Otsu	MST	Proposed
Accuracy(%)	72.26	64.71	80.09

TABLE II: Segmentation accuracy on the SET1 of Otsu [46], MST [47], and the proposed segmentation method.

Fig. 6 shows the segmentation results for two MM lesions. These are difficult cases where one of the algorithms fails to localize the lesions ROI (for example, Otsu in Fig. 6(b), MST in Fig. 6(c)). The melanoma lesion shown in Fig. 6(a), Fig. 6(c) and Fig. 6(e) contains several regions with different visual features. This image was well segmented only by the Otsu method (Fig. 6(a)), while MST was trapped in one of the regions of different color intensity (Fig. 6(c)).

For the MM displayed in Fig. 6(b), Fig. 6(d), and Fig. 6(f) the MST method was able to accurately localize its center part (Fig. 6(d)) while Otsu method fails entirely (Fig. 6(b)). However, by using the proposed method we can determine the lesion boundary when one of the segmentation methods fails.

There are several computation and memory efficient segmentation methods proposed in the literature which may be suitable for mobile-based applications [59]–[61]. Here we give qualitative comparison between the proposed segmentation method with the state-of-the-art level set segmentation method [59] and interactive segmentation method [61].

Comparison with level set segmentation: Level set segmentation is a fast and accurate segmentation method that has been widely used in many applications. Here we compare with a state-of-the-art level set segmentation proposed in [59] and we use the software by the authors for comparison [62]. We found that there are several advantages of our proposed combination of Otsu and MST when comparing with level set segmentation for this application: 1) Level set segmentation would not work well when the boundary of the mole is not clear and the background in the image is complex. As an example from our dataset, Fig. 7, in this case the boundary of the mole is not clear and the background is not uniform, and the level set segmentation boundary is not accurate compared to our proposed method. 2) There are some parameters needed to be tuned for level set segmentation. As there are different types of images in our dataset (i.e., different background, shape, color of moles), it is difficult to find an optimal set of parameters for our dataset and in general for this application. 3) Our proposed method is time efficient and it is faster than level set. Given the image shown in Fig. 7 for example, it takes 0.4 seconds with our method, and around 4 seconds to run with level set in 200 iterations (both experiments are done on the same desktop and implementation in Matlab). Note that level set segmentation requires on an initial rectangle provided by the user to crop around the mole region, and its accuracy depends on this initial rectangle. Thus, it is non-trivial to set up a fair and thorough comparison between level set segmentation and our proposed automatic segmentation. We will investigate this in future work.

Comparison with interactive segmentation: We have compared with a state-of-the-art interactive method [61] for segmentation on our dataset. The code of interactive segmentation is provided by the authors, available from [63].

Some comparison results between interactive and our proposed method are shown in Fig. 8. As we can see from the example images, the segmentation results of interactive method depend significantly on the initial strokes drawn by users. In some cases, even when the initial strokes have only small mistakes, the interactive segmentation results could be quite inaccurate as shown in Fig. 8. In order to properly segment the mole, it requires accurate user input, which maybe sometimes difficult for users, especially elders. Therefore, here we apply automatic segmentation method for our application. The use of interactive segmentation for this application will be investigated thoroughly in future. It is worth noting that the accuracy of interactive segmentation depends on the initial strokes drawn by users. Therefore, it is non-trivial to set up a fair and thorough comparison between interactive segmentation [61] and our proposed automatic segmentation.

C. Feature Selection Results

Our proposed feature selection tool is employed on individual feature categories, i.e., color (54 features), border (16 features) and texture (8 GLCM features and 1 edge density feature). Note that we have not applied the feature selection for the asymmetry category since it contains only one feature. We also have not applied feature selection to LBP, as the LBP descriptors have high dimension and this requires a large dataset to estimate the probability distribution in (3) to compute the mutual information.

In order to compute the mutual information, features should be first discretized. To discretize each feature, the original interval is divided into a number of equal bins. Feature values falling outside the interval are assigned to the extreme left or right bin. We run feature selection with the number of bins equal to $\{2, 3, 4, 5, 6\}$ and let feature selection determine the optimal number of bins. The best classification accuracy is achieved when the number of bins is equal to 5.

The values of n_i^o and n_i^e in (5) are set to 50% number of samples of class containing the sample i , and α in (6) is set to 0.4. For the CT feature the number of partitions considered in feature selection are $PA = \{4, 8, 12, 16\}$, while the number of subparts are $SP = \{2, 4, 8\}$. For the Border Fitting feature the number of lines analyzed are $nt = \{8, 12, 16, 20, 24, 28\}$.

Table III shows selected features in each category when MI-based criterion (2) and our proposed criterion (6) are used during the feature selection [28]. The classification accuracy of different feature categories for MI-based criterion and our proposed criterion is given in Table IV.

From Table III, we can see that CT and Border Fitting features are always selected for both MI and proposed criteria. This confirms the efficiency of our proposed novel features. For the texture category, both feature selection-based MI and the proposed method select same feature set, i.e., *edge density*, *the contrast of 64 quantization level GLCM*, and *correlation of 32 quantization level GLCM*. In following sections, we refer these three selected features as *GLCM+edge* features for distinguishing with LBP_s features.

Additionally, Table IV shows that the proposed feature selection method can select more discriminative sets of features. In particular, for the color feature category, the proposed

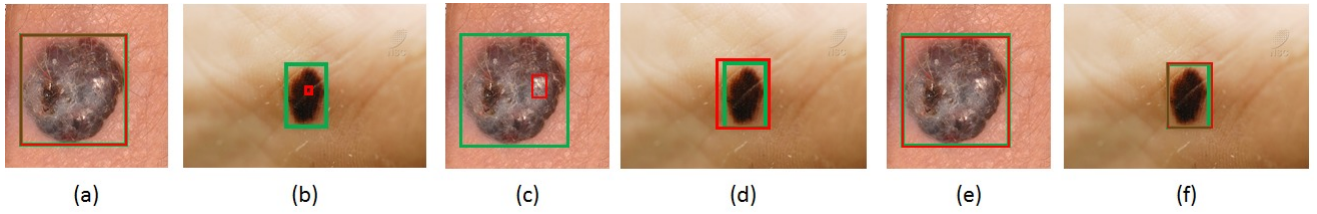


Fig. 6: Segmentation evaluation for the Otsu (a), (b), the MST (c), (d), and the proposed (e), (f) methods. The green rectangle represents the ground-truth, the red rectangle denotes our segmentation result.

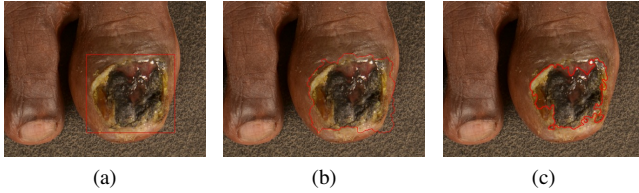


Fig. 7: Comparison with level set segmentation for an example image with complex backgrounds. (a): cropped 0-level initialization for level set segmentation (provided by the user). (b): level set segmentation result. (c): segmentation result using our proposed method.

Feature Category	Mutual Information	Proposed
color	color triangle ($SP = 16, PA = 8$), num_red, num_hue, num_green	color triangle ($SP = 16, PA = 8$) num_gray, num_hue
border	mean of border fitting ($nt = 12$)	variances of border fitting ($nt = \{8, 12\}$)
texture	edge density contrast of GLCM (64 quantization levels) correlation of GLCM (32 quantization levels)	

TABLE III: Selected features from each category when mutual information criterion and proposed criterion are used.

	Mutual Information			Proposed		
	color	border	texture	color	border	texture
Sens	94.00	82.55	84.55	94.18	79.27	84.55
Spec	86.00	66.00	76.67	90.00	76.00	76.67
Acc	90.00	74.27	80.61	92.09	77.64	80.61

TABLE IV: Feature selection results: the classification performance when MI-based criterion and our criterion are used to select features for individual categories. The classification results are with SVM classifier with 5-folds cross validation on SET2.

method can gain about 2% accuracy improvement by selecting a set of only 3 features instead of a set of 4 features of MI-based method. For the border feature category, the MI-based criterion achieve highest accuracy of 74.27% when only one feature is selected. By using MI-based criterion, we cannot obtain higher accuracy even when more border features are included. On the other hand, the highest accuracy of the proposed criterion is 77.64% when 2 border features are selected.

The importance of using multiple features for each cate-

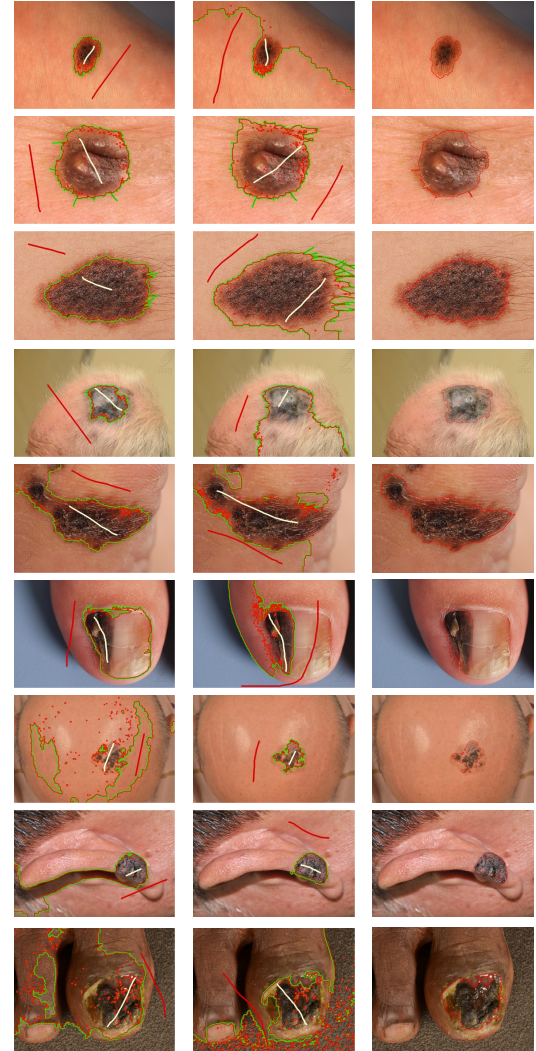


Fig. 8: Example of segmentation result using interactive method [61]. Left & middle columns: Interactive method with different initial strokes (red and white color). Right column: segmentation result using our proposed method.

gory: We remark that we use multiple features for each feature category because of the large variation of the data (e.g., color variation). Using a single feature for each feature category is not enough to distinguish the melanoma and benign samples. For examples, although many melanoma samples usually have the variation in the color (e.g., the color spreads from white to dark), using only *num_gray* feature (the number of non-

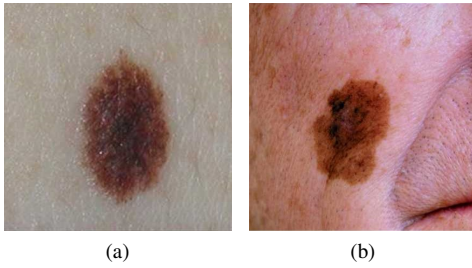


Fig. 9: Examples of (a) benign sample and (b) melanoma sample. Both have high color variation (i.e., high num_gray) values.

Feature	CT	num_gray	num_hue	Combined
Sens	85.21	76.92	88.39	94.18
Spec	77.85	68.97	65.52	90.00
Acc	81.53	72.94	76.96	92.09

TABLE V: The classification performance on the SET2 for each selected color feature (the results are with SVM classifier with 5-folds cross validation). For the Color Triangle (CT) feature, $SP = 16$; $PA = 8$.

Feature	Variances of Border		Combined
	$nt = 8$	$nt = 12$	
Sens	78.20	68.96	79.27
Spec	56.83	57.69	76.00
Acc	67.52	63.32	77.64

TABLE VI: The classification performance on the SET2 for each selected border feature (the results are with SVM classifier with 5-folds cross validation).

zero histogram bins on gray channel) does not give a high classification accuracy. It is because there are some benign samples which also have the color spreading from white to dark. Two examples of the benign (left) and melanoma (right) samples are provided in Fig. 9.

Similarly, there are some benign samples which their color do not vary uniformly from the center to the border. In those cases, using only Color Triangle feature is not sufficient for the classification. Hence, using both num_gray and Color Triangle features helps overcome the data variation issue, and boost the classification accuracy.

We provide here the Table V which shows the classification accuracy, the sensitivity, and the specificity when the selected color features (*Color Triangle* ($SP = 16$; $PA = 8$), num_gray , num_hue) are used individually or are used together. We can see that combining selected color features significantly boosts the accuracy.

We also provide here the Table VI which shows the classification accuracy, the sensitivity, and the specificity when selected border features (*variances of Border Fitting for $nt = 8, 12$*) are used individually or are used together. The results show that combination of selected border features boosts the accuracy.

D. Classification Results

As discussed in section III-D, four binary SVMs are applied to the selected features from four feature categories (color, border, asymmetry, and GLCM+edge) and the k NN is applied for LBP_S features. To aggregate useful information in individual classifiers, these classification results are processed further using several proposed fusion methods.

Performance of individual classifiers: For SVM classifiers, to estimate the generalization error of the models based on the selected features, we employ the 10-folds cross validation on the TEST SET, i.e., 154 segmented ROIs. SVMs return soft values which present the confidence level of a sample belongs to MM. By applying a threshold, we can obtain the hard output values, which is 0 (non-cancer, benign nevus) or 1 (cancer, MM). The columns “Color”, “Border”, “Asymmetry”, “GLCM+edge” of Table VII show the performance of the system on TEST SET for each feature category at the threshold of 0.5.

Fig. 10 shows the visualization of the SVM outputs of the LCF (color features) after dimension reduction and color-coding by the clinical labels. The visualization reveals a good separation of regions corresponding to benign nevus and melanoma classes. This suggests that distinct color characteristics are presences of different skin lesion classes which are captured by the selected features. Particular image samples in Fig. 10 demonstrate how the proposed scheme allows a good separation of skin lesions.

As discussed, for the LBP_S features, we use a k NN classifier. We consider various settings, i.e., parameters for LBP : $R = \{1, 2\}$, $P = \{8, 16\}$; number of neighbors of k NN: $k = \{1, 2, 3\}$. We empirically achieve the best results for following the parameters: $R = 1$, $P = 8$, and $k = 2$. The k NN returns soft values that indicate the confidence level of a sample belonging to MM. Particularly with $k = 2$, k NN returns 1 (or 0) when two nearest neighbors (NN) are MM (or benign); in the remaining situation, i.e., one NN is MM and another NN is benign, the distance from the test sample to its two NNs is utilized to calculate the soft output values. The classification results are showed in the “ LBP_S ” column of Table VII after applying a threshold of 0.5 on the soft output values.

Performance with fusion of individual classifiers: Furthermore, we employ three proposed fusion methods for fusing the results of the four classifiers, i.e., (Color, Border, Asymmetry, and GLCM+edge) or (Color, Border, Asymmetry, and LBP_S), as discussed in section III-D.

For the fusion method 2 (Weighed sum, 2a and 2b), we need to determine the weight values and the threshold to classify cancerous samples. Specifically, by running 3-folds cross validation on SET3, we achieve soft classification values. From these classification values, we calculate the weight, i.e., *sensitivity@50% specificity* and the *Area Under Receive Operating Curve (AUC)*, using for the fusion. Additionally, the threshold is set to the value that maximizes the accuracy on SET3. The weight values are useful in providing a better view of how feature categories contribute to the diagnosis process.

The results showed in Table VIII and ROC in Fig. 11 demonstrate that fusion can improve the overall classification performance. We achieve the best *Sens @ Spec* $\geq 90\%$ of

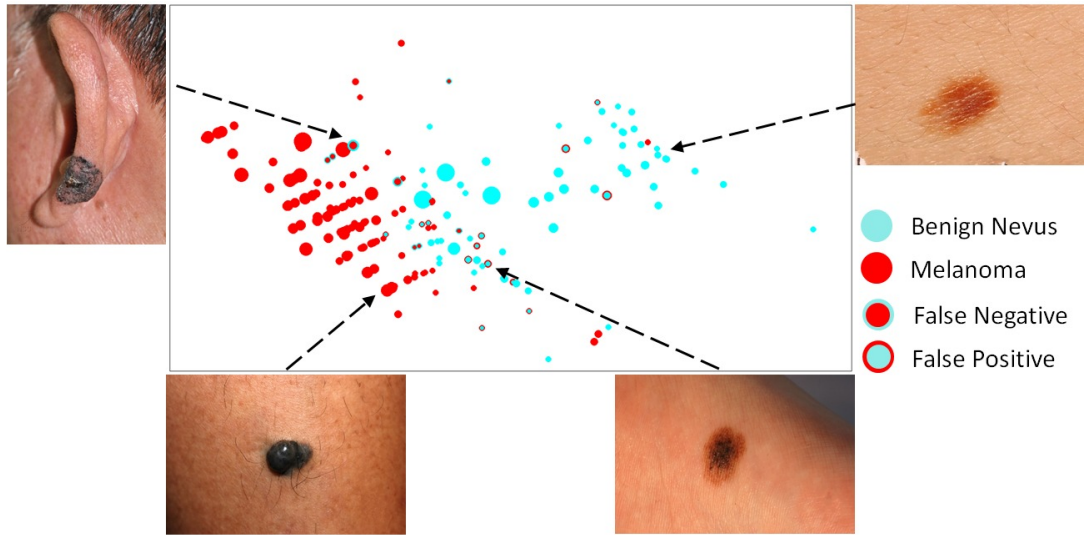


Fig. 10: 2D visualization of SVM output of the LCF after dimension reduction for TEST SET

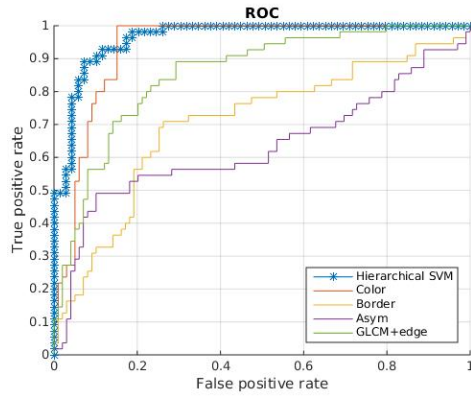


Fig. 11: Receiver Operating Curve (ROC) for four feature categories (color, border, asymmetry, and GLCM+edge) and hierarchical SVM fusion method. (Best view in color)

Feature category	Color	Border	Asymmetry	GLCM+edge	LBP _s
Length	3	2	1	3	36
Sens	96.36	69.09	47.28	69.09	79.10
Spec	83.84	70.71	87.88	82.83	87.17
Acc	90.01	69.90	67.58	75.96	83.14

TABLE VII: Individual Classifiers Performance: The classification performance on the TEST SET for each feature category, i.e., without classifier fusion. The length of each feature category is also provided.

89.09% for the **hierarchical SVM** (color, border, asymmetry, and GLCM+edge) and **AUC weighted sum** (color, border, asymmetry, and LBP_s) methods. The results in Table VIII suggest that different feature categories may not contribute equally in the diagnosis procedure, and proper combination and fusion of them significantly boost the accuracy.

E. Comparison with existing methods

In addition, we compare our proposed melanoma detection system with some recent melanoma detection methods: MED-NODE [31] and [32], on the publicly available UMCG dataset from Department of Dermatology of University Medical Center Groningen (UMCG) [64]. UMCG is a challenging dataset. The dataset consists of 70 melanoma and 100 benign images. MED-NODE [31] performs the training on 45 randomly selected images (20 melanoma and 25 benign images, respectively) and reports the results on the remaining 125 images. [32] performs the training on 125 randomly selected images (50 melanoma and 75 benign images, respectively) and reports the results on the remaining 45 images. For fair comparison, we apply exactly the same data division when comparing to the corresponding methods. It is worth noting that both MED-NODE [31] and [32] use color and texture features to describe the image. MED-NODE [31] reports the classification results for each feature category (i.e., color and texture) independently, while [32] reports the classification results when both color and texture features are combined together. For fair comparison, we also use only our proposed color and texture features in this experiment.

Our system first performs the segmentation (Section III.A), the feature computation for segmented lesion (Section III.B), and the feature selection (Section III.C) on the training dataset. We then report the results with the proposed hierarchical SVM classifier (Section III.D). Similar to [31] and [32], we report following metrics: Sensitivity (Sens), Specificity (Spec), Positive Predictive Value (PPV), Negative Predictive Values (NPV), and Total Accuracy (TAcc - which is the number of correct classified samples over the number of testing samples) to assess the performance of our system. Our results are the average of 50 random splits of training/testing sets. The experimental results presented in Table IX show that our proposed method outperforms recent methods [31] and [32]. It is worth noting that our proposed method is designed to work on mobile device, while it is not clear if MED-NODE [31] and [32] can work on a resource-constrained mobile device.

Feature categories	(Color, Border, Asymmetry, and GLCM+edge)				(Color, Border, Asymmetry, and LBP _s)			
Fusion methods	1	2a	2b	3	1	2a	2b	3
Sens	94.55	94.55	94.55	92.73	100.00	100.00	89.09	98.18
Spec	83.84	74.75	83.84	85.51	84.85	84.85	86.87	82.83
Acc	89.19	84.65	89.19	89.12	92.42	92.42	87.98	90.51
Sens@ Spec $\geq 90\%$	72.73	76.73	87.27	89.09	67.27	73.72	89.09	80.00

TABLE VIII: The performance of the proposed system with fusion of individual classifier results, using different proposed fusion methods: 1, 2a, 2b, and 3 are *Non-weighted sum*, *Sens@ 50%spec weighted sum*, *AUC weighted sum* and *hierarchical SVM* respectively.

	Color		Texture		Color+Texture	
	[31]	Ours	[31]	Ours	[32]	Ours
Sens	74	81	62	66	82	84
Spec	72	73	85	85	71	72
PPV	64	66	74	75	67	70
NPV	81	85	77	79	85	87
TAcc	73	75	76	78	76	77

TABLE IX: The comparative results (in %) between MED-NODE [31] and [32] and our proposed system. MED-NODE [31] reports results for color and texture features independently. [32] reports the results when color and texture features are combined together. The results of MED-NODE [31] and [32] are cited from the corresponding papers. We also follow the experimental setups of the corresponding papers.

Furthermore, we would like to remark that: 1) While there are other mobile applications (Apps) for melanoma detection using smartphones (e.g., SkinVision [65]), we found that they can only take an input image from the phone's camera directly. In particular, these Apps do not allow a user to load a pre-captured image for testing. Therefore, it is not possible to compare their accuracy using a common dataset. 2) Most of these Apps have not been thoroughly tested and their performance is unknown. For a few Apps which have been tested, their systems were evaluated using their own proprietary images [66], making fair comparison with our work difficult. Therefore, we focus on comparison with recent works [31], [32] that have been evaluated using the common dataset, i.e., UMCG [64].

Computational complexity analysis and comparison with recent methods: Several complete systems have been recently proposed for non-dermatoscopic image-based melanoma detection: [31], [32], and the proposed one in this work. Although these systems have similar general stages, i.e., lesion segmentation, feature extraction, and classification, these systems use very different techniques at different stages.

For the lesion segmentation, we use MST and Otsu methods, while [31], [32] used K-means. Given a lesion segmentation, our method uses hierarchical SVM to classify the lesion by using 9 selected features (3 color features, 2 border features, 1 asymmetry feature, and 3 texture features). On the other hand, [32] uses Neural Network (NN) for classification by using 10 color features. [31] combines two different classifiers. The authors use the Cluster-based Adaptive Metric (CLAM) [67] to classify 12 color features. Additionally, the authors also use

Unbiased Color Image Analysis Learning Vector Quantization (Unbiased CIA-LVQ) [68] method to classify texture features.

Here we perform a complexity analysis and comparison at different stages of the processing:

Lesion segmentation stage: Let us define the number of pixels in the skin region as N . For the lesion segmentation, we aim to segment the skin pixels into 2 clusters, i.e., lesion or non-lesion. We use MST and Otsu, and both methods have $\mathcal{O}(N)$ complexity. The works [31], [32] use K-means (in color space) to segment the skin region into two clusters. Hence the complexity of [31], [32] for lesion segmentation step is $\mathcal{O}(NI)$ in which I is the number of iterations of K-means¹. Because I is usually very small in comparison to N , hence at this stage, [31], [32] and ours have comparable complexity.

Feature extraction stage: Let us define the number of pixels in the lesion region as N . For the feature extraction, among our 9 features, Color Triangle (CT) requires the most computation. Specifically, its complexity is $\max(\mathcal{O}(PA^2 \times SP), \mathcal{O}(SP \times N))$, in which PA and SP are the number of parts and subparts of the CT feature. In our experiments, by using $PA = 8$ and $SP = 16$ (Table III) which are much less than the number of pixels in the lesion, hence the complexity of the CT feature is $\mathcal{O}(SP \times N)$.

In [32], the most complex feature is the Color Variation which is computed by Fuzzy C-means (FCM) whose complexity is $\mathcal{O}(NI)$, in which I is the number of iterations of FCM².

In [31], for color features, it uses means and standard deviations of color channels, which have the complexity of $\mathcal{O}(N)$. For texture feature extraction, [31] processes as follows: (i) the salient point detector in [69] is used to detect the 50 most interesting positions (pixels); (ii) at these positions, 15×15 patches are drawn and are used as texture features. Regarding the salient point detection, it is a multi-stage processing, i.e., firstly, Gaussian filters are applied to input at multi-scales to produce multi-scale feature maps. The feature maps are then combined together. Finally, the combined feature map is classified by a neural network to produce salient locations. By using a multi-stage processing which consists of several

¹The general complexity of K-means is $\mathcal{O}(NICd)$, in which N is the number of data points (i.e., pixels); I is the number of iterations of K-means; C is the number of clusters; d is dimension of each data point. Here C and d are very small in comparison to N , i.e., $C = 2$ and $d = 3$, respectively.

²The general complexity of FCM is $\mathcal{O}(NIC^2d)$, in which N is the number of data points (i.e., pixels); I is the number of iterations of FCM; C is the number of clusters; d is dimension of each data point. Here $C = 2$ and $d = 3$, respectively.

complex components, i.e., multi-scale filtering, classification-based neural network, the computational cost of this detector is high, so is the whole process. In summary, for the feature extraction stage, the complexities of our method and [32] are comparable and less than the complexity of [31].

Classification stage: For the classification, our system uses hierarchical SVM comprised of two stages of classifiers. In the first stage, we use 4 RBF-kernel SVM (one for each feature category). In the second (last) stage, another RBF-kernel SVM is used to fuse the soft decisions of the first stage's classifiers. The complexity of a RBF-kernel SVM is $\mathcal{O}(sd)$, where s is the number of support vectors, which is around 50 in our experiments, and d is the number of features. In particular, in the first stage, we use $d = \{3, 2, 1, 3\}$ for color, border, asymmetry, and texture features respectively, and $d = 4$ for the second stage's classifier.

The work [32] uses 2 layer feedforward neural network with the sigmoid activation. Hence, its complexity is $\mathcal{O}(dH)$, in which $d = 10$ is the number of features and $H = 15$ is number of nodes in a hidden layer.

In [31], the complexities of CLAM and Unbiased CIA-LVQ classifiers are $\mathcal{O}(d^2)$, where d is the numbers of features. The input of CLAM is only 12 color features, so it is expected to have a fast inference. However, for Unbiased CIA-LVQ classifier, it first performs the classification for each of 50 patches in which the patch size (feature size) is $d = 15 \times 15$ pixels. It then fuses these 50 outputs to produce final classification result. Hence, the classification stage of [31] is dominated by Unbiased CIA-LVQ classifier which is higher complexity than the classifiers of [32] and ours.

Overall, according to the complexity analysis of the main stages, our method's complexity is comparable to [32] and is smaller than that of [31]. Furthermore, it is worth noting that the processing time of [31], [32] on resource-constrained smartphones has not been examined (the source codes of [31], [32] are not available). On the other hand, for our method, the running time on a smartphone is presented in Section IV-F. Specifically, on the device "Samsung Galaxy S4 Zoom smartphone, with Dual-core CPU running at 1.5GHz Cortex-A9, GPU: Mali-400, RAM: 1.5GB, and storage memory of 8GB", our method takes less than 5 seconds to process an image.

F. Mobile Implementation and Human-Computer Interface (HCI) Design

We implement our proposed image analysis engine and the entire system on a consumer electronic mobile device: Samsung Galaxy S4 Zoom smartphone, with Dual-core CPU running at 1.5GHz Cortex-A9, GPU: Mali-400, RAM: 1.5GB and storage memory of 8GB. The features of the backside camera (the one used during the tests) are: 16 MP, image size: 4608×3456 pixels, with autofocus and $10\times$ optical zoom. We measured the average processing time spent for each image and it is less than 5 seconds. It is worth pointing out that, the mobile phone implementation of the algorithm has not been explicitly parallelized using the available GPU.

HCI Design Study. We conducted a study with an aim to understand the best HCI design principles for this applications. The subjects selected for the study were recruited through a

local community center (i.e., older participants with $\mu=63.0$ and $\sigma=4.76$) and from a local university (i.e., young participants with $\mu=24.1$ and $\sigma=3.22$). The whole evaluation study lasts for 1.5 hours.

When entered the specific test, each participant was required to take a photo of his/her arm skin and then to click the "Start Diagnosis" button. After that, a progress bar was shown to indicate the processing time. The participant was told that he/she could stop the diagnosis process by clicking the "Give up to know the result" button at any point during the processing time. After the session, the researcher debriefed to the participant that the participant should consult professional doctors if he/she is interested in understanding his/her health condition towards the presented disease.

After that, we conducted semi-structured interviews with participants exposed to our prototype. Since our application (a self-diagnosis application) can be considered as a subset of personal informatics systems, we constructed our interview questions based on the structure of stage-based model proposed in [58]. We aimed to identify the design challenges across the entire adoption process. Some important aspects regarding acceptance, collection, integration, reflection, and action, which have emerged from the study are:

- *Acceptance.* One HCI issue that has emerged is the "perceived harmfulness" of the system. Despite our application being implemented on a smartphone, older participants raised questions and concerns about the potential harmfulness of the application. Older adults may internally associate our application with the hospital used diagnostic machines, and they questioned about whether some harmful detection methods (e.g., X-Ray, radiation) were used. Even for the older adults who have been using smartphones for a long period they have similar concerns.
- *Reflection.* Some participants expressed concerns regarding the presentation of the diagnosis results, as P23 mentioned: "I want to know the result, but I don't want the application directly tells me the conclusion. Instead, could it show me the percentage of risk ... or how many people have been diagnosed the disease by this application and the possibility of being proved fine finally?" P5 mentioned: "It is very scary if the machine tells me I have skin cancer ...".
- *Action.* For an effective self-diagnosis application, it is important to improve the trustfulness of the results and encourage users to make certain action (e.g., visit a hospital for a formal check). Some participants expressed the expectation towards the format of the diagnosis result, as P25 mentioned: "I hope the application can provide me more information, such as how the machine gets the diagnosis result. Which features, or what cause the disease happened...otherwise, I couldn't make decisions about the next step."

These HCI issues need to be addressed for successful adoption of the proposed system.

V. DISCUSSION

The prototype presented in this paper has several limitations. First, it is necessary to further validate the system with a

large database. It is worth pointing out that currently there are no publicly available clinical large datasets of camera phone images. It is very challenging and time consuming to generate large datasets (especially for the MM cases), as these images need to be analyzed histopathologically to produce the gold standard.

Second, on some occasions, the images acquired by the mobile devices can be heavily affected by distortions, in particular, motion blur and illumination variations [70], [71]. These distortions can change the skin lesion appearance and smooth the border of the skin mole, confusing the segmentation and classification algorithms that our current system cannot compensate them. It is desirable to include a module to detect such distorted images and alert users in such situations.

Third, the feature categories exploited by our system are extracted merely from the skin images and attempt to incorporate the main dermatologic signs. Our system can be enhanced by incorporating patient-derived clinical data (e.g., age, gender, size and location of the lesion). We believe that this side information will improve the sensitivity and specificity of the proposed system.

VI. CONCLUSION

We propose an accessible mobile health-care solution for melanoma detection, using mobile image analysis. The main characteristics of the proposed system are: an efficient hierarchical segmentation scheme suitable for the resource-constrained platform, a new set of features which efficiently capture the color variation and border irregularity from the smartphone-captured image, and a new mechanism for selecting a compact set of the most discriminative features. The experimental results based on 184 camera images demonstrate the efficiency of the prototype in accurate segmentation and classification of the skin lesion in camera images. We foresee several possible usage scenarios for the current solution: it could be employed by the general public for preliminary self-screening or it can assist the general physicians during the diagnosis process.

In addition to the technical development, we paid attention also to understand the usability and acceptance challenges. For this purpose, we have investigated the HCI design issues through an exploratory case study and semi-structured interviews. Our study discovered several important HCI issues that should be addressed in future work.

ACKNOWLEDGMENT

The authors would like to thank Dr. Martin Chio and Dr. Michelle Liang from NSC for providing us clinical opinions and the dataset, and helping to validate the system. Furthermore, we would like to thank Prof. Susan Swetter, from Stanford University Medical Center and Cancer Institute, for her valuable comments during this research. The material reported in this article is supported in part by the MIT-SUTD International Design Centre (IDC). Any findings, conclusions, recommendations, or opinions expressed are ours and do not necessary reflect the views of the IDC.

REFERENCES

[1] W. Zhou, M. Yang, H. Li, X. Wang, Y. Lin, and Q. Tian, "Towards codebook-free: Scalable cascaded hashing for mobile image search," *IEEE Transactions on Multimedia*, pp. 601–611, April 2014.

[2] H. Li, Y. Wang, T. Mei, J. Wang, and S. Li, "Interactive multimodal visual search on mobile device," *IEEE Transactions on Multimedia*, vol. 15, no. 3, pp. 594–607, April 2013.

[3] B. Girod, V. Chandrasekhar, D. M. Chen, N.-M. Cheung, R. Grzeszczuk, Y. Reznik, G. Takacs, S. S. Tsai, and R. Vedantham, "Mobile visual search," *IEEE Signal Processing Magazine*, vol. 28, no. 4, pp. 61–76, 2011.

[4] T. Chen, K. H. Yap, and D. Zhang, "Discriminative soft bag-of-visual phrase for mobile landmark recognition," *IEEE Transactions on Multimedia*, vol. 16, no. 3, pp. 612–622, April 2014.

[5] W. Min, C. Xu, M. Xu, X. Xiao, and B. K. Bao, "Mobile landmark search with 3d models," *IEEE Transactions on Multimedia*, vol. 16, no. 3, pp. 623–636, April 2014.

[6] F. Cricri, M. J. Roininen, J. Leppänen, S. Mate, I. D. D. Curcio, S. Uhlmann, and M. Gabbouj, "Sport type classification of mobile videos," *IEEE Transactions on Multimedia*, vol. 16, no. 4, pp. 917–932, June 2014.

[7] P. Bao and D. Gourlay, "A framework for remote rendering of 3-D scenes on limited mobile devices," *IEEE Transactions on Multimedia*, vol. 8, no. 2, pp. 382–389, April 2006.

[8] F. Liu, Y. Zhang, S. Liu, B. Zhang, Q. Liu, Y. Yang, J. Luo, B. Shan, and J. Bai, "Monitoring of tumor response to au nanorod-indocyanine green conjugates mediated therapy with fluorescence imaging and positron emission tomography," *IEEE Transactions on Multimedia*, vol. 15, no. 5, pp. 1025–1030, Aug 2013.

[9] M. X. Ribeiro, A. J. M. Traina, C. T. Jr., and P. M. Azevedo-Marques, "An association rule-based method to support medical image diagnosis with efficiency," *IEEE Transactions on Multimedia*, vol. 10, no. 2, pp. 277–285, Feb 2008.

[10] A. M. Alaa, K. H. Moon, W. Hsu, and M. van der Schaar, "Confident-care: A clinical decision support system for personalized breast cancer screening," *IEEE Transactions on Multimedia*, vol. 18, no. 10, pp. 1942–1955, Oct 2016.

[11] M. E. Celebi, T. Mendonca, and J. S. Marques, *Dermoscopy Image Analysis*. CRC Press, 2015.

[12] H. Nejati, V. Pomponiu, T.-T. Do, Y. Zhou, S. Irvani, and N.-M. Cheung, "Smartphone and mobile image processing for assisted living," *IEEE Signal Processing Magazine*, vol. 33, no. 4, pp. 30–48, 2016.

[13] J. Wolf, J. Moreau, O. Akilov, and et al., "Diagnostic inaccuracy of smartphone applications for melanoma detection," *JAMA Dermatology*, vol. 149, no. 4, pp. 422–426, 2013.

[14] L. Rosado, M. J. M. Vasconcelos, R. Castro, and J. M. R. S. Tavares, *Dermoscopy Image Analysis. Chapter 12: From Dermoscopy to Mobile Teledermatology*. CRC Press, 2015.

[15] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2017," *CA: A Cancer Journal for Clinicians*, vol. 67, no. 1, pp. 7–30, 2017.

[16] A. A. Marghoob and A. Scope, "The complexity of diagnosing melanoma," *J Invest Dermatol*, vol. 129, no. 1, pp. 11–13, 2009.

[17] A. R. Sadri, M. Zekri, S. Sadri, N. Gheissari, M. Mokhtari, and F. Kolahdouzan, "Segmentation of dermoscopy images using wavelet networks," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 4, pp. 1134–1141, Apr. 2013.

[18] American Academy of Dermatology, "Melanoma: Signs and symptoms," <https://www.aad.org/public/diseases/skin-cancer/melanoma>, Accessed on May 13, 2017.

[19] G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco, and M. Delfino, "Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions," *Archives of Dermatology*, vol. 134, no. 12, Dec. 1998.

[20] S. W. Menzies, C. Ingvar, and W. H. McCarthy, "A sensitivity and specificity analysis of the surface microscopy features of invasive melanoma," *Melanoma Research*, vol. 6, no. 1, pp. 55–62, 1996.

[21] U. D. of Health and H. Services, "Mobile medical applications. guidance for industry and food and drug administration staff," <http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm089643.htm>, may 2013.

[22] I. Maglogiannis and C. Doukas, "Overview of advanced computer vision systems for skin lesions characterization," *IEEE Trans. Inform. Technol. Biomed.*, vol. 13, no. 5, pp. 721–733, Sep. 2009.

[23] H. Ganster, A. Pinz, R. Röhner, E. Wildling, M. Binder, and H. Kittler, "Automated melanoma recognition," *IEEE Trans. Med. Imaging*, vol. 20, no. 3, pp. 233–239, 2001.

[24] M. E. Celebi, H. A. Kingravi, B. Uddin, H. Iyatomi, Y. A. Aslandogan, W. V. Stoecker, and R. H. Moss, "A methodological approach to the classification of dermoscopy images," *Comp. Med. Imag. and Graph.*, vol. 31, no. 6, pp. 362–373, 2007.

- [25] T. Wadhawan, N. Situ, H. Rui, K. Lancaster, X. Yuan, and G. Zouridakis, "Implementation of the 7-point checklist for melanoma detection on smart handheld devices," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011.
- [26] K. Ramlakhan and Y. Shang, "A mobile automated skin lesion classification system," in *ICTAI*, 2011.
- [27] D. Charalampos, S. Paris, K. C. T. and M. Ilias, "Automated skin lesion assessment using mobile technologies and cloud platforms," in *Conf Proc IEEE Eng Med Biol Soc*, 2012, pp. 2444–2447.
- [28] T.-T. Do, Y. Zhou, H. Zheng, N.-M. Cheung, and D. Koh, "Early melanoma diagnosis with mobile imaging," in *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014.
- [29] T.-T. Do, Y. Zhou, V. Pomponiu, N.-M. Cheung, and D. C. I. Koh, "Method and device for analysing an image," Patent US 20 170 231 550 A1, August 17, 2017.
- [30] Z. Guo, D. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1657–1663, 2010.
- [31] I. Giotis, N. Molders, S. Land, M. Biehl, M. F. Jonkman, and N. Petkov, "Med-node: A computer-assisted melanoma diagnosis system using non-dermoscopic images," *Expert Systems with Applications*, vol. 42, no. 19, pp. 6578 – 6585, 2015.
- [32] M. H. Jafari, S. Samavi, S. M. R. Soroushmehr, H. Mohaghegh, N. Karimi, and K. Najarian, "Set of descriptors for skin cancer diagnosis using non-dermoscopic color images," in *IEEE International Conference on Image Processing (ICIP)*, Sept 2016.
- [33] R. Garnavi, M. Aldeen, and J. Bailey, "Computer-aided diagnosis of melanoma using border- and wavelet-based texture analysis," *IEEE Trans. Inform. Technol. Biomed.*, vol. 16, no. 6, pp. 1239–1252, Nov. 2012.
- [34] S. Haider, D. Cho, R. Amelard, A. Wong, and D. A. Clausi, "Enhanced classification of malignant melanoma lesions via the integration of physiological features from dermatological photographs," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014.
- [35] M. Emre Celebi, Q. Wen, S. Hwang, H. Iyatomi, and G. Schaefer, "Lesion border detection in dermoscopy images using ensembles of thresholding methods," *Skin Research and Technology*, vol. 19, no. 1, pp. 252–258, 2013.
- [36] E. L. Torre, T. Tommasi, B. Caputo, and G. E. Gigante, "Kernel methods for melanoma recognition," in *MIE*, 2006, pp. 983–988.
- [37] S. Demyanov, R. Chakravorty, M. Abedini, A. Halpern, and R. Garnavi, "Classification of dermoscopy patterns using deep convolutional neural networks," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, April 2016, pp. 364–368.
- [38] J. Kawahara, A. BenTaieb, and G. Hamarneh, "Deep features to classify skin lesions," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, April 2016, pp. 1397–1400.
- [39] Y. Gu and J. Tang, "A mobile system for skin cancer diagnosis and monitoring," *SPIE Mobile Multimedia/Image Processing, Security, and Applications*, May 2014.
- [40] O. Abuzaghlleh, M. Faezipour, and B. D. Barkana, "Skincure: An innovative smart phone-based application to assist in melanoma early detection and prevention," *CoRR*, vol. abs/1501.01075, 2015.
- [41] T. T. K. Munia, M. N. Alam, J. Neubert, and R. Fazel-Rezai, "Automatic diagnosis of melanoma using linear and nonlinear features from digital image," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July 2017.
- [42] M. H. Jafari, E. Nasr-Esfahani, N. Karimi, S. M. R. Soroushmehr, S. Samavi, and K. Najarian, "Extraction of skin lesions from non-dermoscopic images for surgical excision of melanoma," *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, no. 6, pp. 1021–1030, Jun 2017.
- [43] V. Pomponiu, H. Nejati, and N.-M. Cheung, "Deepmole: Deep neural networks for skin mole lesion classification," in *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016.
- [44] 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*, vol. 542, 2017.
- [45] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection," *IJCV*, vol. 46, no. 1, pp. 81–96, Jan. 2002.
- [46] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, Jan 1979.
- [47] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *IJCV*, vol. 59, no. 2, pp. 167–181, 2004.
- [48] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, Nov 1973.
- [49] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, pp. 51–59, 1996.
- [50] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.
- [51] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *TPAMI*, vol. 27, pp. 1226–1238, 2005.
- [52] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Transactions on Neural Networks*, vol. 5, pp. 537–550, 1994.
- [53] P. A. Estévez, M. Tesmer, C. A. Perez, and J. M. Zurada, "Normalized mutual information feature selection," *IEEE Transactions on Neural Networks*, vol. 20, no. 2, pp. 189–201, Feb 2009.
- [54] F. Wang, X. Wang, D. Zhang, C. Zhang, and T. Li, "marginface: A novel face recognition method by average neighborhood margin maximization," *Pattern Recognition*, vol. 42, no. 11, pp. 2863–2875, 2009.
- [55] K. Fukunaga, *Introduction to Statistical Pattern Recognition (2nd Ed.)*. San Diego, CA, USA: Academic Press Professional, Inc., 1990.
- [56] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM TIST*, vol. 2, no. 3, pp. 1–27, 2011.
- [57] L. Nanni, A. Lumini, and S. Brahnam, "Local binary patterns variants as texture descriptors for medical image analysis," *Artificial Intelligence in Medicine*, vol. 49, no. 2, pp. 117–125, Jun. 2010.
- [58] I. Li, A. Dey, and J. Forlizzi, "A stage-based model of personal informatics systems," in *Annual ACM Conference on Human Factors in Computing Systems (CHI)*, 2010.
- [59] C. Li, C. Xu, C. Gui, and M. D. Fox, "Distance regularized level set evolution and its application to image segmentation," *IEEE Transactions on Image Processing*, vol. 19, no. 12, 2010.
- [60] Y. Zhou, T.-T. Do, H. Zheng, N.-M. Cheung, and L. Fang, "Computation and memory efficient image segmentation," *IEEE Transactions on Circuits and Systems for Video Technology*, 2016.
- [61] V. Gulshan, C. Rother, A. Criminisi, A. Blake, and A. Zisserman, "Geodesic star convexity for interactive image segmentation," in *CVPR*, 2010.
- [62] C. Li, C. Xu, C. Gui, and M. D. Fox, "Matlab implementation of Level Set for Image Segmentation," <https://www.mathworks.com/matlabcentral/fileexchange/12711-level-set-for-image-segmentation>, 2013.
- [63] V. Gulshan, C. Rother, A. Criminisi, A. Blake, and A. Zisserman, "Matlab implementation of Geodesic Star Convexity for Interactive Image Segmentation," https://github.com/gavin-s-smith/Floor_ID/tree/master/gsc-1.2, 2012.
- [64] [Online]. Available: http://www.cs.rug.nl/~imaging/databases/melanoma_naevi/
- [65] "SkinVision," <https://skinvision.com/>.
- [66] T. Maier, D. Kulichova, K. Schotten, R. Astrid, T. Ruzicka, C. Berking, and A. Udrea, "Accuracy of a smartphone application using fractal image analysis of pigmented moles compared to clinical diagnosis and histological result," *Journal of European Academy of Dermatology and Venereology*, vol. 29, pp. 663–667, 2015.
- [67] "Cluster-based adaptive metric classification," *Neurocomputing*, vol. 81, pp. 33 – 40, 2012.
- [68] I. Giotis, K. Bunte, N. Petkov, and M. Biehl, "Adaptive matrices and filters for color texture classification," *Journal of Mathematical Imaging and Vision*, vol. 47, no. 1, pp. 79–92, Sep 2013.
- [69] D. Walther and C. Koch, "Modeling attention to salient proto-objects," *Neural Networks*, vol. 19, no. 9, pp. 1395 – 1407, 2006.
- [70] F. Xie, Y. Lu, A. C. Bovik, Z. Jiang, and R. Meng, "Application-driven no-reference quality assessment for dermoscopy images with multiple distortions," *IEEE Transactions on Biomedical Engineering*, vol. 63, pp. 1248–1256, 2016.
- [71] Y. Lu, F. Xie, Y. Wu, Z. Jiang, and R. Meng, "No reference uneven illumination assessment for dermoscopy images," *IEEE Signal Processing Letters*, vol. 22, pp. 534–538, 2015.