Identification of Skin Cancer based on Colour, Subregion and Texture

Alin Mariam Solomon¹, Anju Murali², Sruthi R B³, Sreekavya M K⁴, Saranya Sasidharan⁵, Leena Thomas⁶
B.Tech Graduate¹, 2, 3, 4, 5, Assistant Professor⁶
Department of ECE
College of Engineering, Kottarakara, India

Abstract:
Dermatology is a medical discipline of analysis and treatment of skin anomalies. Skin lesions are a major public health problem nowadays because of various known and unknown factors. The early diagnosis of skin disease is very important for a good prognosis and complete cure. From the dermatological skin image the border between lesion and surrounding skin is detected. The features like colour, texture, shape, diameter are checked by the gray level co-occurrence matrix and further classified using layered model and flat model. Thus the lesion can be classified as melanoma, basal cell carcinoma, squamous cell carcinoma and nevus.

Keywords: dermatology, melanoma, basal cell carcinoma, squamous cell carcinoma, gray level co-occurrence matrix, layered model

I. INTRODUCTION
Incidence of skin cancer has been increasing over the decades with early detection and treatment becoming more and more important. Melanoma, the most fatal of all skin cancers has a survival rate of only 9-15% at stage IV while this rate increases to 85–99% if detected early at stage II. Melanomas typically occur in the skin but may rarely occur in the mouth, intestines, or eyes. Basal Cell Carcinoma (BCC) is the most common skin cancer. It is rarely life threatening but if left untreated it can erode the surrounding tissues and can cause disfigurement of the affected part. Thus, early detection and excision is essential for the treatment of BCC. Nevus, also known as a mole, is the medical term for sharply circumscribed and chronic lesions of the skin or mucosa. A Seborrheic Keratosis (SK) is a benign skin growth that originates in keratinocytes. The lesions appear in various colors, from light tan to black.

The purpose of this paper is to distinguish among four types of skin lesions: melanoma, nevus, BCC, and SK, using a significantly larger dataset. Melanoma and BCC account for 80% of all skin cancer incidences. Accurate identification of nevus and SK are clinically important since they are sometimes confused with melanomas.

Skin is the largest organ of the body and protects the body from injury and infection. It also regulates the body temperature and plays a key role in protecting the body from germs. Skin consists of three layers. They are epidermis, dermis and hypodermis. The top layer is the epidermis and it is made up of three types of cells. The flat cells at the top of this layer are called squamous cells. The cells under the squamous cell are basal cells. The cells found in the lower part of epidermis are melanocytes and they are also responsible for skin color pigmentation. Skin cancer can be divided into two types. They are malignant and benign. Malignant skin cancer develops from melanocytes and they are also called as melanoma. They can originate in any part of the body that contains melanocytes. They are much less common than benign skin cancer. Benign skin cancer can be divided into basal cell carcinoma (BCC) which develops from basal cells and squamous cell carcinoma (SCC) which develops from squamous cells. Malignant melanoma is nowadays one of the leading cancers among many white-skinned populations around the world. Change of recreational behaviour together with the increase in ultraviolet radiation cause a dramatic increase in the number of melanomas diagnosed. It arises from cancerous growth in pigmented skin lesion. It is a slowly progressive condition which begins in the melanocytes in skin. Melanocytes are the pigmentation cells of the skin.

II. METHODOLOGY USED
This project works on the development of classification methods for both MSLs and NoMSLs. First, we developed a general border detection algorithm for MSLs and NoMSLs.

Figure 1: Block Diagram
Finding the border of NoMSLs was a challenging task because they often have unclear borders. After determining the border of the tumor, segmented the skin lesion image into four regions as normal skin, peripheral, central tumor, and whole tumor. It is
again classified into three steps: color related features, sub region related features, texture related features.

III. METHOD PROPOSED

A. BORDER DETECTION

From each skin lesion image, extract the border between the tumor and the surrounding normal skin area. Accurate border detection usually results in better classification performance. The core of the algorithm is color thresholding, removal of artifacts such as microscope border and hair, and inclusion of bright area seen specifically in NoMSLs. First, we developed a general border detection algorithm for MSLs and NoMSLs. Finding the border of NoMSLs was a challenging task because they often have unclear borders. With this sophisticated algorithm, we found that a linear classifier with only two image features (“skewness of bright region on the major axis” and “difference in blue intensity between the peripheral and the normal skin”) discriminated MSLs from NoMSLs with performance of 98.0% SE and 86.6% SP. The project focused on further development of the system to detect melanomas from other MSLs (nevi) and NoMSLs.

B. FEATURE EXTRACTION

After determining the border of the tumor, segment the skin lesion image into four regions as normal skin, peripheral, central tumor, and whole tumor. The whole tumor consists of all pixels within the extracted border. In contrast, the normal skin is all pixels on the outside of the border. The peripheral is the first 30% of the whole tumor area, obtained by going inward from the border. Finally, the central tumor is obtained by removing the peripheral from the whole tumor. The features are grouped into the three categories: color, sub-region, and texture.

1 COLOR RELATED FEATURES

As for color related feature, calculated 5 statistics (min, max, standard deviation, skewness, entropy) of the intensity of six color channels (R: red, G: green, B: blue, H: hue, S: saturation, and V: luminance) for each of the three tumor regions (peripheral, central tumor, and whole tumor s). This yielded 90 parameters (5 statistics × 6 channels × 3 regions). Then calculated the difference in the same five statistics on the six color channels between central tumor and peripheral and those between peripheral and normal skin area, which yielded 60 parameters (5 statistics × 6 channels × 2 pairs-of-regions). These difference oriented features to be robust over variations of dermoscopy images caused by different photographic conditions. In total, there are 150 color-related features (90+60).

II SUB-REGION RELATED FEATURES

Sub-region related features describe geometrical distribution of the color. Then divided the central tumor and the peripheral into smaller even sub-regions. Here two types of subdivisions are used: angle wise and distance wise. The angle wise is based on the angle from the center of gravity of the central tumor to the edge of the region. The distance wise is based on the Euclidean distance from the outer border. Three numbers of sub-regions used are: 4, 8, and 16 for the angle-wise manner and 2, 4, and 8 for the distance-wise manner.

For each sub-region, calculated three statistics (mean, standard deviation, and skewness) on four color channels (R, G, B, and S). This left out H and V because these two channels did not contribute to the classification performance in our preliminary experiments. Finally, it calculated the standard deviation of these statistics within all sub-regions. This yielded 144 sub-region features (2 target regions × 2 types of subdivisions × 3 numbers of sub-regions × 4 color channels × 3 statistics for each sub-region). In the previous studies had the asymmetry features to describe For each sub-region, calculated three statistics (mean, standard deviation, and skewness) on four color channels (R, G, B, and S). This left out H and V because these two channels did not contribute to the classification performance in our preliminary experiments. Finally, it calculated the standard deviation of these statistics within all sub-regions. This yielded 144 sub-region features (2 target regions × 2 types of subdivisions × 3 numbers of sub-regions × 4 color channels × 3 statistics for each sub-region.

III TEXTURE RELATED FEATURES

As for texture related features, we adopted the gray level co-occurrence matrix (GLCM). Then obtained the GLCMs with the following settings: two target regions (central tumor and whole tumor), three quantization levels (N = 16, 32, and 64), four distances (δ = 1, 2, 4, and 8 pixels), and four directions (θ = 0°, 45°, 90°, and 135° from the major axis). From each GLCM, we extracted four GLCM statistics (energy, correlation, entropy, and homogeneity).

To make the directional settings (θ) more meaningful extracted min, mean, max, and difference (i.e., max−min) of the aforementioned GLCM-statistics in four main directions (θ) as was also recommended in the original literature of the GLCM. In total, there are 384 texture features (2 regions × 3 quantization levels × 4 distances × 4 directions (e.g., max) × 4 GLCM statistics).

C. CLASSIFICATION

This section introduces the layered model as the primary classification model and the flat model as a performance baseline shown in Fig.3.3 and 3.4, respectively. The letters M, N, B, and S in the figures denote melanoma, nevus, BCC, and SK, respectively. Linear classifiers over nonlinear ones are used to gain a clear understanding of the relationship between the inputs and the outputs of the models and to facilitate a comparison of the classification performance.

LAYERED MODEL

The first-step classifier “MN-BS” identifies the input skin lesion as MSL if the output value is greater than the classifier’s threshold value or as NoMSL otherwise. These are shown by (+) and (−) in Fig.3.3. If the result is an MSL, the second-step classifier “M-N” distinguishes melanoma from nevus in the same manner by comparing its output value with the threshold value. If the result from the first-step classifier is a NoMSL, the second-step classifier “B-S” distinguishes BCC from SK.

The idea of the layered model is to decompose the whole classification task to 1) the broad classification of MSL and NoMSL by the “MN-BS” and 2) the detailed classification of “melanoma and nevus” and that of “BCC and SK” by the “MN” and “B-S,” respectively. It may be inferred that the first-step
classifier “MN-BS” must have high accuracy because misclassifications at this phase are fatal. We designed this model based on the results of our past studies that distinguishing MSLs from NoMSLs is relatively easy. One of the most important steps for the classifier development is feature selection. It is well known that too many features or irrelevant features lead to poor performance and the overfitting problem. Therefore, it is necessary to select an appropriate subset of features for each of the classifiers “MN-BS,” “M-N,” and “B-S.”

FLAT MODEL
This method introduce two types of flat models, namely the ‘flat model I” and the ‘flat model II” as the performance baseline. Each of the flat models has four linear classifiers: “M,” “N,” “B,” and “S” whose output values estimate the presence/absence of the corresponding classes: melanoma, nevus, BCC, and SK, respectively. This kind of classification model is typically used for the multiclass classification.

The flat model I and the flat model II are different in how the classifiers possess the features. In the flat model I, all classifiers share the same features. The features are selected with the Wilks’ Lambda stepwise method with the strategy that it improves overall classification performance. In the flat model II, each classifier possesses its own features. Then select the features specifically effective for each classifier by the Wilks' Lambda stepwise method as well as the layered model.

The screenshot of the images are shown in the figure. The input image is first read out. It is then undergone pre processing and lesion is separated from skin and all the colour, subregion and texture features are studied and based on that the cancer is detected.

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

REFERENCES
[1] M. Emre Celebi, Senior Member, IEEE, and Azaria Zornberg, “Automated Quantification of Clinically Significant Colors in Dermoscopy Images and Its Application to Skin Lesion Classification”, IEEE SYSTEMS JOURNAL, VOL. 8, NO. 3, SEPTEMBER 2014
