Hierarchical Classification of Ten Skin Lesion Classes

Hierarchical Classification of Ten Skin Lesion Classes

Carlo Di Leo, Vitoantonio Bevilacqua — Department of Electric and Information Engineering, Polytechnic of Bari, Bari,Italy

Lucia Ballerini, Robert Fisher — School of Informatics, University of Edinburgh, Edinburgh, UK

Ben Aldridge, Jonathan Rees — Dermatology, University of Edinburgh, Edinburgh, UK

This paper presents a hierarchical classification system based on the k-Nearest Neighbors (kNN) classifier for classification of ten different classes of Malignant and Benign skin lesions from color image data.

Our key contribution is to focus on the ten most common classes of skin lesions. There are five malignant: Actinic Keratosis (AK), Basal Cell Carcinoma (BCC), Squamous Cell Carcinoma (SCC), Melanoma (MEL), Intraepithelial Carcinoma (IEC) and five benign: Melanocytic Nevus / Mole (ML), Seborrhoeic Keratosis (SK), Dermatofibroma (DF), Haemangioma (VASC), Pyogenic Granuloma (PYO). Moreover, we use only high resolution color images acquired using a standard camera (non-dermoscopy).

Our image dataset contains 1300 lesions belonging to ten classes (45 AK, 239 BCC, 331 ML, 88 SCC, 257 SK, 76 MEL, 65 DF, 97 VASC, 24 PYO and 78 IEC).

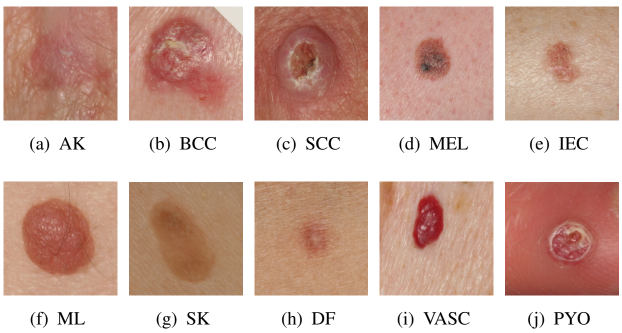

Figure 1: Examples of skin lesion images from the different classes used in this work

The ground truth used for the experiments is based on agreed classifications by two dermatologists and a pathologist. Images are acquired using a Canon EOS 350D SLR camera. Lighting was controlled using a ring flash and all images were captured at the same distance ($\sim$50 cm) resulting in a pixel resolution of about 0.03 mm. Lesions are segmented using the method described in [3]. Specular highlights have been removed [1].

The classification approach used in our research falls within the hierarchical model. Our approach divides the classification task into a set of smaller classification problems corresponding to the splits in the classification hierarchy. Each of these subtasks is simpler than the original task, since the classifier at a node in the hierarchy need only distinguish between a smaller number of classes. Therefore, it may be possible to separate the smaller number of classes with higher accuracy and more specialized properties. Moreover, it may be possible to make this determination based on a smaller set of features. The hierarchy is fixed *a priori* by grouping image classes into two main groups. The first group contains malignant and pre-malignant lesion classes (AK, BCC, SCC, MEL, IEC). The second group contains benign lesions and some Melanoma (ML, SK, MEL, DF, VASC, PYO). The two groups were constructed to enhance classification between malignant and benign lesions but all lesions inside a group have similar characteristics: benign lesions are brown or reddish and are characterized by uniform color and a regular shape, while malignant lesions have pink, white, red and brown colors, high color variation within the same lesion, and random shapes. While all non-melanoma skin lesions are well classified at the root node, some MEL have the same brown color as ML and SK whose samples in the dataset are an order of magnitude more common than MEL samples. This means that half of MEL were misclassified as Benign lesions and this is why, to overcome this problem, Melanoma lesions are classified in both of the two groups. Thus, if a Melanoma is classified as a benign lesion by the first classifier, it is still possible to correct the misclassification. Each of the 3 decision nodes uses a k-Nearest Neighbor Classifier (kNN).

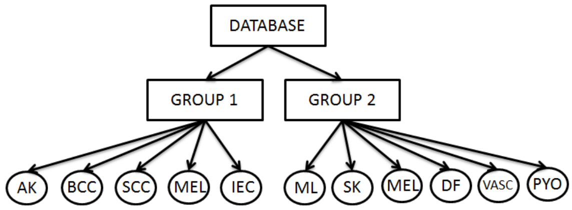

Figure 2: Hierarchical organisation of our skin lesion classes

In this work three different measures have been combined to calculate the distance between lesions: the Bhattacharyya Distance, the Histogram Intersection Distance and the Euclidean Distance. The features used in the classification system are widely used features extracted from shape, color and texture information [2]. The Bhattacharyya Distance has been used for comparing the lesion's mean color and covariance matrix [1]. The color histogram provides information about the image's color distribution. It is invariant to image rotation and translation and tolerant to changes in the image's scale. A 3D color histogram uses the values of the three color channels. Features using the Euclidean Distance come from six different families: Texture features (extracted from the grayscale co-occurrence matrix of color channel pairs as given in [1]); Shape-related Features (Perimeter-Area Ratio, Form Factor, Ratio between the skin lesion area and the area of the minimum bounding box, Eccentricity); Asymmetry; Fourier Descriptors; ad hoc Color Ratio Features; Invariant Moments; Histogram-based Descriptors [2].

Sequential forward feature selection (SFS) was applied for each classifier in the hierarchical system. SFS finds the features that maximize classification accuracy for a given level.

Accuracy is 93% when discriminating malignant from benign lesions and it reaches an overall classification accuracy of 67% over ten classes of skin lesions, five malignant and five benign.

	Malignant	**Benign**
Malignant	**491 (93%)**	35 (7%)
Benign	115 (15%)	**659 (85%)**

Table 1: Confusion matrix on distinguishing between malignant and benign classes: rows are true classes, columns are the selected classes. The classification rate is inserted within brackets

[1] Lucia Ballerini, Robert B. Fisher, Ben Aldridge, and Jonathan Rees. A color and texture based hierarchical k-nn approach to the classification of non-melanoma skin lesions. In M. Emre Celebi and Gerald Schaefer, editors, *Color Medical Image Analysis*, volume 6 of *Lecture Notes in Computational Vision and Biomechanics*, pages 63–86. Springer Netherlands, 2013. ISBN 978-94-007-5388-4.

[2] Carlo Di Leo. Design of a content-based medical image retrieval and classification system for ten skin lesions classes. Master's thesis, Polytechnic of Bari, 2013. URL `http://dx.doi.org/10.13140/2.1.4067.6644`.

[3] Xiang Li, Ben Aldridge, Lucia Ballerini, Robert Fisher, and Jonathan Rees. Depth data improves skin lesion segmentation. In Guang-Zhong Yang, David Hawkes, Daniel Rueckert, Alison Noble, and Chris Taylor, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2009)*, volume 5762 of *Lecture Notes in Computer Science*, pages 1100–1107. Springer Berlin Heidelberg, 2009. ISBN 978-3-642-04270-6.