2011

BCC Skin Cancer Diagnosis Based on Texture Analysis Techniques

Shao-Hui Chuang
Old Dominion University

Xiaoyan Sun
Old Dominion University

Wen-Yu Chang
Kaohsiung Medical University

Gwo-Shing Chen
Kaohsiung Medical University Hospital

Adam Huang
National Central University Jhongli

See next page for additional authors

Original Publication Citation

Chuang, S. H., Sun, X., Chang, W. Y., Chen, G. S., Huang, A., Li, J., & McKenzie, F. D. (2011). BCC skin cancer diagnosis based on texture analysis techniques. In R.M. Summers & B.V. Ginneken (Eds.), *Medical Imaging 2011: Computer-Aided Diagnosis, Proceedings of SPIE Vol. 7963* (796330). SPIE of Bellingham, WA. https://doi.org/10.1117/12.878124

BCC SKIN CANCER DIAGNOSIS BASED ON TEXTURE ANALYSIS TECHNIQUES

Shao-Hui Chuang[1], Xiaoyan Sun[1], Wen-Yu Chang[2], Gwo-Shing Chen[3], Adam Huang[4], Jiang Li[1] and Frederic D. McKenzie[1]
[1]Department of Electrical and Computer Engineering, Old Dominion University, Norfolk, VA, USA
[2]Graduate Institute of Medicine,
[3]Department of Dermatology, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan
[4]Research Center for Adaptive Data Analysis, National Central University, Jhongli, Taiwan

ABSTRACT

In this paper, we present a texture analysis based method for diagnosing the Basal Cell Carcinoma (BCC) skin cancer using optical images taken from the suspicious skin regions. We first extracted the Run Length Matrix and Haralick texture features from the images and used a feature selection algorithm to identify the most effective feature set for the diagnosis. We then utilized a Multi-Layer Perceptron (MLP) classifier to classify the images to BCC or normal cases. Experiments showed that detecting BCC cancer based on optical images is feasible. The best sensitivity and specificity we achieved on our data set were 94% and 95%, respectively.

Keywords: BCC Skin Cancer, Texture Analysis, Gray Level Run Length Matrix (GLCM), Gray Level Co-occurrence Matrix (GLCM)

1. INTRODUCTION

Skin cancer is the most common form of cancer in the United States. More than one million skin cancers are diagnosed annually [1]. Comparing to other types of cancers, skin cancer can be detected easily because the tumors usually develop in the outermost layer of skin [2]. Dermatologists detect skin cancer through a visual examination of the skin. If a suspicious malignancy is found, a biopsy will be performed. Basal Cell Carcinoma (BCC) is a non-melanoma skin cancer. Research is usually focused on classifying microscopy skin samples into different cancer categories [3]. However, a microscopy sample is required excisional biopsy or incisinonal biopsy procedures, which are uncomfortable to patients. Our purpose is to develop a computer-aided diagnosis (CAD) system that not only provides a second opinion to dermatologists, but also avoids unnecessary lesion removal if it is not malignancy. We proposed a CAD system that differentiates BCC skin cancer from two benign skin growths, the Seborrheic Keratoses (SK) and Nevus based on a texture analysis technique. The gray level Run Length Matrix (GLRLM) and gray level Co-covariance Matrix (GLCM) based texture features were first extracted from each individual image. Those features then went through a feature selection procedure and finally, the selected features were utilized by a MLP classifier to classify each of the images into one of the categories (BCC or normal).

2. BACKBROUND

The GLRLM method was proposed by Galloway in 1975 [4] and by Chu et al. in 1990 [5]. By definition, consecutive pixels with the same gray level, in a given direction, constitute a gray level run, and the number of pixels in the run is called run length. The GLCM method was proposed by Haralick in 1973 [6]. These two methods require two parameters which are distance and angles. Once the matrices are calculated along each direction (0°, 45°, 90° and 135°), those feature descriptors can be used either with respect to each direction or a combination of all directions to get a global view of texture information.

Feature selection is used to select a useful feature subset by eliminating features with little or no predictive power or features that are highly correlated with already chosen features. A piecewise linear network method was used to perform feature selection in this paper [7]. The algorithm selects the most important features by performing the orthonormal least square procedure through a piecewise linear network (PLN) model for the given data set.

Medical Imaging 2011: Computer-Aided Diagnosis, edited by Ronald M. Summers, Bram van Ginneken,
Proc. of SPIE Vol. 7963, 79633O · © 2011 SPIE · CCC code: 0277-786X/11/$18 · doi: 10.1117/12.878124

K-fold cross validation (CV) is used to evaluate our CAD system for the given data set. K-fold cross validation is one of the most popular evaluation methods for classification, where the data set is divided into k subsets, and one subset is used for testing and the remaining subsets are used for training. This procedure is repeated k times so that each of the subsets is used for testing once. Performance is then computed based on the testing results. Three-fold CV is used in this paper.

Multi-Layer Perceptron (MLP) classifier is one of the commonly used neural network classifier. The MLP classifier is capable of learning a rich variety of nonlinear decision surfaces. MLP is a feedforward artificial neural network model that is designed through iterative regression to minimize the mean squared error (MSE) between the target class and a network mapping function. The three-layer MLP (input-hidden-output) is the most frequently used structure.

A confusion matrix table is the easiest way to represent the diagnosis results. A true positive indicates that a positive image is correctly diagnosed as positive and a false positive denotes that a normal case is wrongly identified as positive. Similarly, a false negative indicates that the classifier misses a positive image and a true negative implies that the classifier correctly classifies normal cases as normal. Sensitivity is the ratio of the number of true positives over the actual true positives. Specificity is the ratio of the number of true negatives over the total number of actual normal images, and accuracy is the sum of true positives and true negatives over the total number of images.

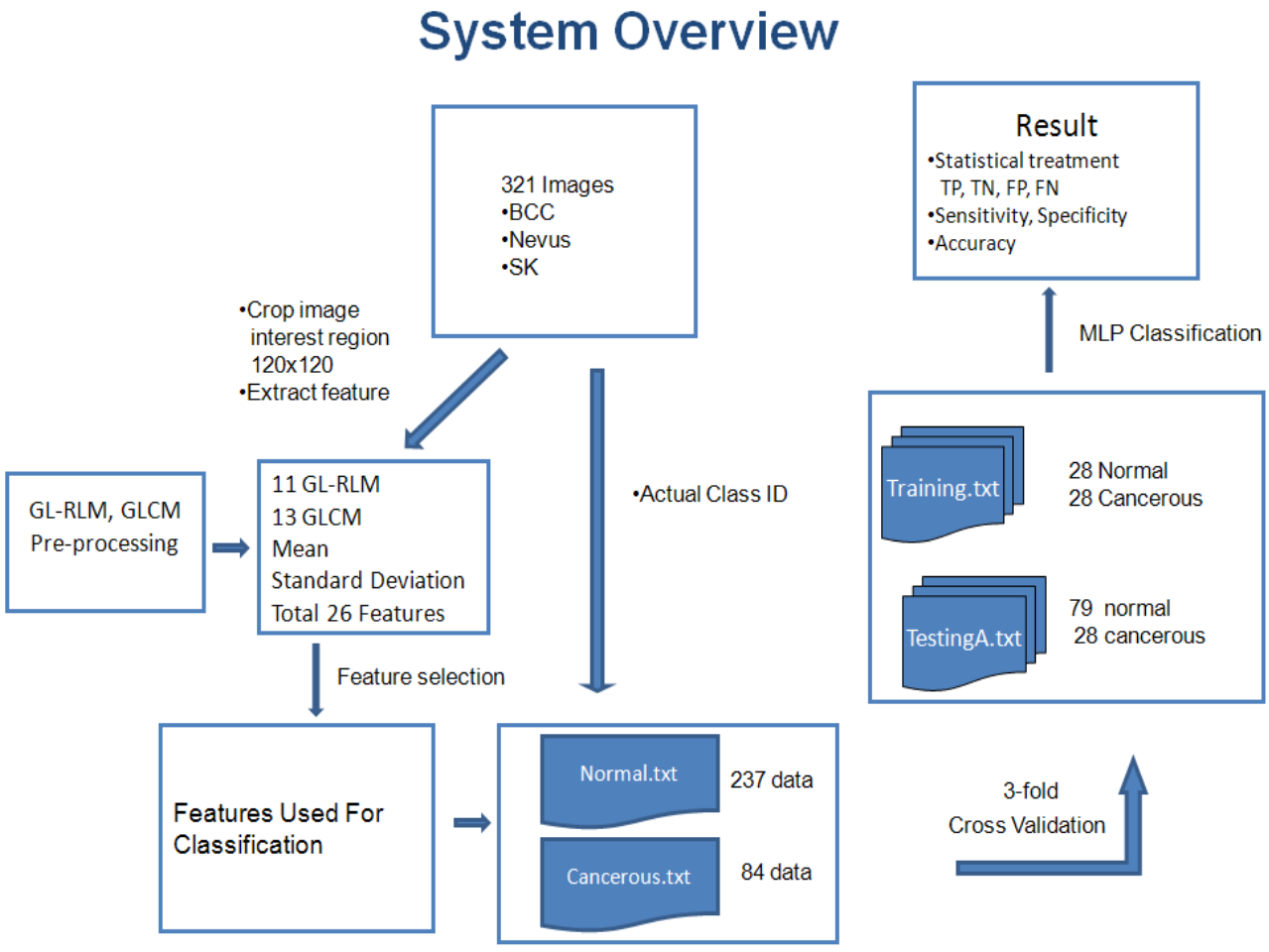

Figure 1. System Overview

3. METHODS

The developed system consists of three components: image preprocessing, feature selection and classification (Figure 1). First, RGB color images from suspicious skin regions taken by regular camera were converted to gray level images. The original images were then cropped into a small rectangle sub-image that only contains the interest region with various sizes. In general, BCC has a larger size than that of SK and nevus. To avoid this size variability, we performed three experiments using different preprocessed image sizes: original, normalized and center region images. The normalized images were generated by normalizing each image to the average size in one dimension. This can avoid providing prior size information for feature extraction. However, normalized images may provide extra information for extracting RLM features because resized images affect the number of run length. The images cropped in the center of interest regions with a size of 120x120 pixels did not go through the resizing step.

Second, we extracted texture features from the three sets of images. There were 26 features extracted from each image: 13 features from GLCM, 11 features from GLRLM. The given images were initially categorized into 3 types: BCC, Nevus and SK. In this paper, we focus on separating BCC from Nevus and SK, where BCC was considered as cancer while Nevus and SK were treated as normal.

Finally, we applied the feature selection algorithm [7] to identify a useful feature set and applied the 3-fold cross validation method using a MLP classifier to evaluate our CAD systems. Figure 1 shows the system overview.

4. EXPERIMENTS AND RESULTS

4.1 Data acquisition

The data set used in this paper was previously collected at the Kaohsiung Medical University Hospital in Taiwan. There are 321 images in total, in which 84 images are for BCC cancer, 235 images are for Nevus and SK benign lesions. Each of the images was confirmed by a medical doctor through a biopsy microscopy procedure.

For each of the image sets (original, normalized and center images), we performed three experiments in which the MLP classifier used different feature sets for the diagnosis: the GLRLM features, the GLCM features and combined GLRLM and GLCM features.

4.2 Original Images

First, for the GLCM feature sets, the feature selection algorithm selected 10 features including Information Measures of Correlation 1, Entropy Difference Entropy, Angular Second Moment, Sum Average, Inverse Difference Moment, Correlation, Information Measures of Correlation 2, Sum Variance and Sum Entropy. Table 1 shows the 3-fold cross validation results. An average sensitivity of 72.62±0.012% and an average specificity of 61.60±0.047% were achieved. The average accuracy was 64.49±0.0337%.

Second, if we use the GLRLM features for the diagnosis, the feature selection algorithm chose 5 useful features for the classification including Short Run High Gray-Level Emphasis, Gray-Level Non-Uniformity, Run Length Non-Uniformity, Low Gray-Level Run Emphasis and Short Run Emphasis. The sensitivity was 89.29±0.02%, the specificity was 98.73±0.004% and the accuracy was 96.26±0.003%.

Third, the feature selection algorithm identified 16 features from 26 combined GLRLM and GLCM features for the classification including Run Length Non-uniformity, Short Run Low Gray-Level Emphasis, Long Run Emphasis, Gray Level Non-uniformity, Short Run High Gray-Level Emphasis, Long Run High Gray-Level Emphasis, Short Run Emphasis, Long Run Low Gray-Level Emphasis, Contrast, High-gray-Level Run Emphasis, Low Gray-Level Run Emphasis, Information Measure of Correlation 1, Correlation, Difference Entropy, Inverse Difference Moment and Run Percentage. The sensitivity in this experiment was 84.13±0.050%, the specificity was 96.61±0.008% and the accuracy

was 93.35±0.010%. Table 1 shows results from the three experiments which indicate that GLRLM features are more effective for the diagnosis.

Table 1. Results for Original Images

	GLCM	GLRLM	COMBINED
Features Used	12,9,11,1,6,5,3, 13,7,8	6,9,10,3,1	R10, R5, R2, R9, R6, R8, R1, R7, F2, R4, R3, F12, F3, F11, F5, R11
Sensitivity	72.62±0.01%	89.29±0.02%	84.13±0.05%
Specificity	61.60±0.05%	98.73±0.00%	96.61±0.01%
Accuracy	64.49±0.03%	96.26±0.00%	93.35±0.01%

4.3 Normalized Images

We performed the same experiments on this data set as that on the normalized images. Table 2 shows selected features, sensitivities and specificities in each of the experiments. Again, the GLCM features performed better and the combined features improved the classification significantly.

Table 2. Results for Normalized Images

	GLCM	GLRLM	COMBINED
Feature Used	12,6,7,13,5,4,8	5,1,9,13,3,4,12, 7	R11,R5,F5,R2,R7,F3, F12,F10,R6,R9,R1,R4
Sensitivity	75.40±0.04%	91.67±0.04%	94.44±0.02%
Specificity	76.09±0.02%	95.08±0.01%	95.22±0.01%
Accuracy	75.91±0.01%	94.19±0.01%	95.02±0.01%

4.4 Center Region Images

Table 3 shows classification results obtained from the center region images. It is observed that the Run Length Matrix based features can provide a better classification.

Compare with these results from the three sets of images, classification based on the normalized images has the highest accuracy. Short Run Emphasis, Gray-Level Non-uniformity, Run Length Non-uniformity, Low Gray-Level Run Emphasis and Short Run High Gray-Level Emphasis from GLRLM features are the most common selected features. Those features are most important features to distinguish BCC cancer from benign lesion.

5. DISCUSSION

Figure 2 shows two true positive examples (BCC cancer) which were correctly classified as cancer by the MLP classifier. Figure 3 shows two examples of false negative that were missed by the classifier. It is observed that true positive images have darker gray level than false negative images do. Bleeding is also one of the characteristics for BCC images. It is found retrospectively that the two false negatives were images taken from patients who have their BCC cancer being removed so that the appearance of the suspicious region was significantly changed.

Table 3. Results for 120x120 Pixels Images

	GLCM	GLRLM	COMBINED
Feature Used	4,5,6,7,3,13,11,1,12	3,12,1,10,11,5,8,4,9,2,7	R9,F9,F7,F11,R6,R3,F2,R12,F5,R1,R2,F13,F3,R7,R11,F12
Sensitivity	63.89±0.03%	90.08±0.04%	78.81±0.04%
Specificity	61.60±0.03%	97.47±0.00%	92.83±0.02%
Accuracy	62.20±0.01%	95.54±0.01%	87.85±0.01%

Figure 4 shows examples of false positive that the actual condition is a normal image and the system diagnosed wrongly as a BCC image. Figure 5 shows examples of true negative images that are either a Nevus or a SK lesion. It is observed that false positive images show smooth edges and shinny surface of tumor than true negative images do. The surfaces of true negative images are ragged.

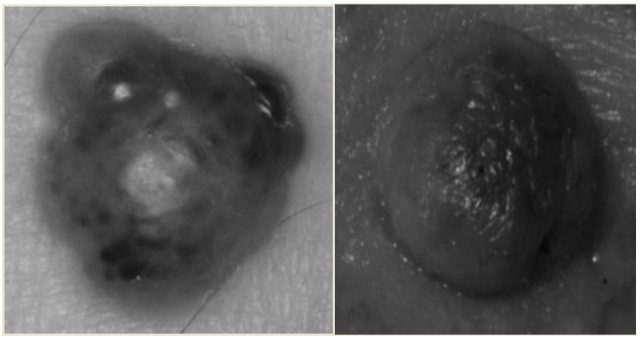

Figure 2. Examples of True Positive Images

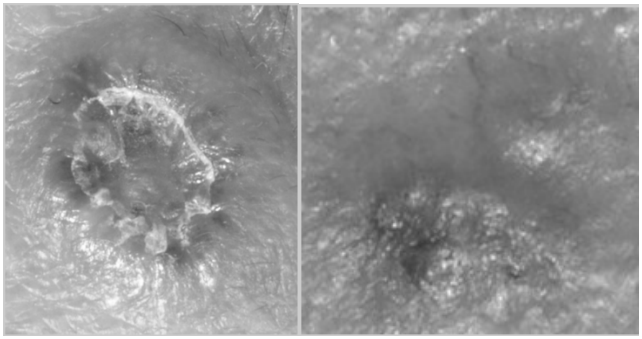

Figure 3. Examples of False Negative Images

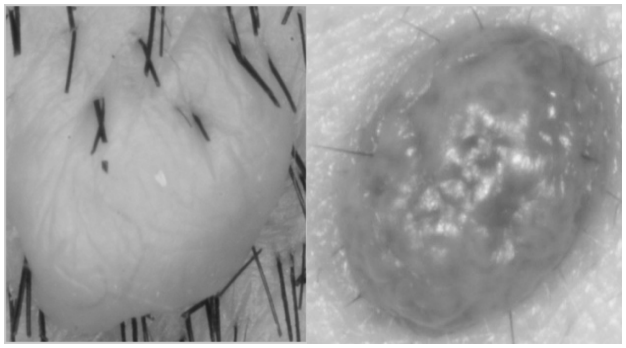

Figure 4. Examples of False Positive of Images

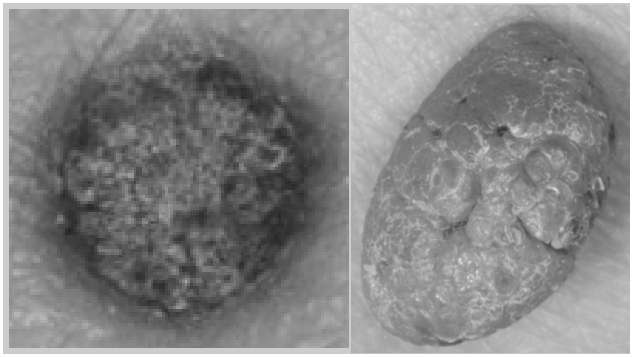

Figure 5. Examples of True Negative of Images

6. CONCLUSION AND FUTURE WORK

We successfully developed a system that can categorize BCC, SK and Nevus skin lesion images taken by a regular optical camera. Our experiments showed that GLRLM and GLCM features are useful to diagnose BCC shin cancer. In particular, the GLRLM features including Short Run Emphasis, Gray-Level Non-uniformity, Run Length Non-uniformity, Low Gray-Level Run Emphasis and Short Run High Gray-Level are the most effective features for the Diagnosis. Image normalization also plays an important role in the diagnosis. Our future work includes collecting more data to future evaluate our CAD system and improve our system to characterize different skin lesions such as Nevus and SK. The current feature sets are derived from gray level images, it might be beneficial to investigate features in color spaces [8].

7. REFERENCE

[1]. American Cancer Society. *Cancer Facts & Figures 2009*. Atlanta: American Cancer Society; 2009.

[2]. National Cancer Institute. "Skin Cancer". U.S. National Institutes of Health. 2009. http://www.cancer.gov/cancertopics/types/skin

[3]. S. Segura, S. Puig, C. Carrera, J. Palou, J. Malvehy, "Development of a two-step method for the diagnosis of melanoma by reflectance confocal microscopy." J. Am Acad Dermatol. vol. 61, no. 2, pp. 216-29, 2009.

[4]. M. M. Galloway, "Texture Analysis Using Gray Level Run Lengths", Computer Graphics and Image Processing, pp. 172-179, 1975.

[5]. A. Chu, C. M. Sehgal and J.F. Greenleaf, "Use of Gray Value Distribution of Run Lengths for Texture Analysis", Pattern Recognition Letters, vol. 11, no. 6, pp.415-419, 1990.

[6]. R.M. Haralick, K. Shanmugam et al., "Textural Feature for Image Classification", IEEE Transactions on systems, man, and cybernetics, vol. 3, no. 6, pp. 610-621, 1973.

[7]. J. Li, M. T. Manry, et al., "Feature Selection Using a Piecewise Linear Network", IEEE Transaction on networks, vol. 17, no. 5, pp. 1101-1115, 2006.

[8]. J. Kong, O. Sertel, H. Shimada, K.L. Boyer, J. H. Saltz and M.N. Gurcan, "Computer-aided Evaluation of Neuroblastoma on Whole-slide Histology Images: Classifying Grade of Neuroblastic Differentiation", Pattern Recognition, vol. 42, no. 6, pp. 1080-1092, 2009.