Computer based diagnosis of thyroid cancer based on cytological images

Abstract

The purpose of the present thesis is the study of classification algorithms for the design of a pattern recognition system to characterize thyroid gland cancer to benign or malignant categories, based on the processing and analysis of cytological images.

The process includes the digitization of microscopy images from the cytological material, in specially prepared specimens, processed with Hematoxylin-Eosin stain. The material includes twenty (20) thyroid gland biopsy samples from corresponding cases, diagnosed by an experienced histopathologist physician.

For the design and the implementation of the system, ten (10) benign and ten (10) malignant images of the thyroid gland were used. Initially, the system embraces pre-processing and segmentation algorithms of cytological images for finding the regions of interest (nuclei). Specifically, segmentation algorithms for microscopy images were studied (thresholding with global threshold, thresholding with adaptive threshold, Otsu method) for finding the optimal solution and were compared with corresponding commercial packages, such as ImPro. The resulted images of this package were distinctive at very large degree, without overlapping of cellular nuclei. Based on ImPro segmentation, 217 segmented nuclei were obtained from benign images and 328 segmented nuclei were found from malignant images. Next, five (5) morphological features and twelve (12) textural features were extracted from the nuclei regions. Then, the Minimum Distance, the Nearest Neighbor, the Bayesian and the Probabilistic Neural Network classifiers were implemented for the classification of nuclei in two classes. For each classifier, the optimal combination of features was found, using the Sequential Backward Selection and the Exhaustive Search as feature selection methods. The proposed system was evaluated by Leave-One-Out method. The accuracy of the system in 'new' data was evaluated by External Cross Validation method.

Eight (8) features were presented statistically significant differences ($p<0.001$) according to Wilcoxon statistical test. The system classified with 95% overall accuracy the nuclei of the thyroid gland in the two classes (benign/malignant), using the 3 Nearest Neighbor classifier. The proposed system is capable to classify 'an unkown' nuclei with 93% accuracy.

SUBJECT AREA: Digital image processing and analysis

KEYWORDS: image segmentation, textural features, feature selection, pattern recognition, thyroid cancer, light microscopy, cytological image