

Classification of pre-malignant lesions of cervical cancer based on microscopy images

Abstract

The purpose of this thesis is the design of a pattern recognition system, which will be able to classify pre-cancerous lesions of the cervix in low (CIN1) and high risk (CINII, CINIII) by processing histopathological images, so that the treatment of patients can be optimized.

More specifically, biopsies of thirty five (35) patients were used, which were diagnosed with intra-epithelial lesions. Of the total samples, twelve (12) were diagnosed as low-grade risk (CIN1), while the remaining twenty three (23) were diagnosed as high-grade risk (CIN2 and CIN3) in accordance with the diagnosis of an experienced histopathologist, who used the system of classification of the World Health Organization.

Initially, we had to prepare the samples, by staining the tissue samples with H&E (Hematoxylin & Eosin) and then the images were digitized. From each sample we digitized images, which were taken from the representative regions (Region of Interest(ROI)), that were focused by the histopathologist. The segmentation of the nuclei is the next step, which was carried out by using the method of thresholding and the morphological operator of opening. Then, seventeen (17) features were extracted from the segmented nuclei, from which five (5) were morphological and twelve (12) textural features.

These features acted as input in a pattern recognition system that was designed to diagnose the type of the intraepithelial lesion. This system was constructed by the classifiers: «k-Nearest Neighbors (k-NN)», «Linear Bayes Classifier», «Probabilistic Neural Network (PNN)», «Minimum Distance Classifier (MDC)», «Least Squares Minimum Distance Classifier (LSMDC)», «Support Vector Machines (SVM)» and «Single Perceptron». The system was used for classifying the nuclei into the two categories (low or high risk).

For each classifier, the optimal feature combination was found applying the Sequential Backward Selection and the Exhaustive Search feature selection methods. The system was evaluated using the Leave-One-Out (LOO) method. The system evaluation in “new” data was performed by means of the External Cross Validation (ECV) method.

Seven of the examined features the skewness, the kurtosis, the contrast, the correlation, the short run emphasis, the gray level non uniformity and the run length non uniformity showed statistically significant differences ($p < 0.001$) between the two categories when the Wilcoxon non-parametric statistical test was employed.

The proposed pattern recognition system discriminates the pre-cancerous image lesions of the cervix in low (CIN I) and high risk (CIN II, CIN III) categories with 94.6%-95% overall accuracy using the PNN classifier and the ECV evaluation method. The features that have the highest incidence in best feature combination are the area, the kurtosis, the eccentricity and the gray level non uniformity. These characteristics describe the heterogeneity and morphology in the image.

SUBJECT AREA: Digital image processing and analysis

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