Postgraduate student: **Theodosi Angeliki**

Thesis Title:

Image analysis and classification of 2D gel electrophoresis images based on pattern recognition methods

Abstract:

The aim of the present study was the design and development of a computer based system for the classification of two dimensional gel electrophoresis images, from patients with either Myeloid or Lymphoid Leukemia, using image analysis methods and pattern recognition techniques.

A significant advantage of two-dimensional polyacrylamide gel electrophoresis of proteins images (2D-gel electrophoresis), is the plethora of proteins presented on a single gel, where as at the same time it may be difficulty to detect protein-spots of high significance (possible biomarkers). This paper presents the design and development of an image processing and analysis system for the detection of 2D gels spots that can effectively distinguish between patients with Myeloid Leukemia (ML) and Lymphoid Leukemia (LL). The publicly available LECB 2-D PAGE gel images database was employed. These images were produced by scanned 2D electrophoresis gels and the sites of the particular spots of interest were known.

Regions of Interest (ROIs) were extracted both manually and automatically by defining an area of 13x13 pixels, the center of which was stored in the database as centroid of the spot. Thus, two sets of ROIs were obtained, one manually and one automatically. For each spot a number of textural features were calculated based on first and second order statistical measures.

Feature reduction took place by applying the Man Whitney test for the detection of statistical significant differences ($p<0.001$). Thus, for each one of the 22 spots, three features were retained having the higher statistical significant differences between the two categories (ML and LL). Then, the features were ranked based on the combination criterion of correlation and Wilcoxon statistical test.

Consequently, based on the retained feature vectors, a Probabilistic Neural Network (PNN) classifier and a kNN classifier were trained so as to discriminate the two classes (ML-LL). Feature selection was based on the exhaustive search method. The system classification accuracy was evaluated by the Leave One Out (LOO) method, and the External Cross Validation method (ECV).

The overall accuracy of the system, when the PNN classifier and the LOO method were employed was found to be 98.1%. Regarding the ECV method, the mean accuracy was 89.4±4.5%. Concerning the kNN classifier and the LOO method the overall accuracy was 96.3% for with k=3, while employing the ECV evaluation method an average of 89.5±4.9% was attained for k=3.

Regarding the automatically obtained ROIS, results were 96.3% for the PNN classifier and the LOO method, and when the PNN was evaluated by the ECV method the overall accuracy was reduced to 88.6±3.6%. Employing the LOO method for the evaluation of kNN classifier with

k=3, the overall accuracy was 97.2%, whereas employing the ECV method the mean overall classification accuracy was 91.1±4.6%.
Features extracted from spots {8, 12, 22}, as defined in the original online database, seem to have significant differentiating capabilities between the ML and LL 2D-gel images. Those features were the average and range of the Energy, the average and range of the Contrast, and the average of Homogeneity.

SUBJECT AREA: Image Processing

KEYWORDS: 2D Gel Electrophoresis, Myeloid Leukemia, Lymphoid Leukemia, Pattern Recognition, Biomarkers, PNN

Examining Committee:

Phd. Dionisis Cavouras, Professor, Department of Medical Instruments Technology, TEI-A (thesis advisor)

Phd. Manolis Sangkriotis, Associate Professor, Dept. of Informatics and Telecommunications, UoA

Phd. Erricos Ventouras, Professor, Department of Medical Instruments Technology, TEI-A