



**ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ**

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ  
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

**ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ  
"ΤΕΧΝΟΛΟΓΙΕΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΣΤΗΝ ΙΑΤΡΙΚΗ ΚΑΙ ΤΗ ΒΙΟΛΟΓΙΑ"**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

**Σύστημα αυτόματης διάγνωσης ιστοπαθολογικών  
εικόνων μικροσκοπίας.**

**Λούης Παπαγεωργίου**

**Επιβλέπων: Κάβουρας Διονύσιος , Καθηγητής**

**ΑΘΗΝΑ**

**Ιούνιος 2013**

## **ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

Σύστημα αυτόματης διάγνωσης ιστοπαθολογικών  
εικόνων μικροσκοπίας

**Λούης Παπαγεωργίου**

**Επιβλέπων:** **Κάβουρας Διονύσιος**, Καθηγητής  
**A.M.:** ΠΙΒ059

**ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ:** **Κάβουρας Διονύσιος**, Καθηγητής  
**Σαγκριώτης Μανώλης**, Καθηγητής  
**Σπύρος Κωστόπουλος**, Εργαστηριακός Συνεργάτης

Ιούνιος 2013

## ΠΕΡΙΛΗΨΗ

Αντικείμενο της παρούσας διπλωματικής είναι η μελέτη και ανάπτυξη ενός συστήματος λήψης, επεξεργασίας και ανάλυσης ιστοπαθολογικών εικόνων μικροσκοπίας για την υποστήριξη της διάγνωσης του ενδομητριακού καρκίνου. Στην υπάρχουσα βιβλιογραφία του ενδομητριακού καρκίνου, δεν υπάρχει άλλη μελέτη που να αφορά την ποσοτική ανάλυση ιστοπαθολογικών εικόνων ενδομητριακών κυττάρων. Παρόμοιες εργασίες, αφορούσαν ολοκληρωμένα συστήματα ταξινόμησης ιστοπαθολογικών εικόνων άλλου είδους καρκίνου. Η δομή της υλοποίησης χωρίζεται σε πέντε βασικές διαδικασίες.

Η πρώτη διαδικασία αφορά την ψηφιοποίηση των διαθέσιμων βιοψιών του ενδομητριακού καρκίνου. Οι βιοψίες επεξεργάστηκαν από την ιστοπαθολόγο με τη χρώση *Αιματοξυλίνης-Ιωσίνης* για τη εκτίμηση του βαθμού διαφοροποίησης του όγκου και ανοσοϊστοχημικά για την έκφραση του ογκογονιδίου *cerb-B*. Έχουμε ψηφιοποιήσει 17 βιοψίες/περιστατικά, μέσα από τις οποίες έγινε λήψη σε 117 ψηφιακές ιστοπαθολογικές εικόνες μικροσκοπίας. Η λήψη των ιστοπαθολογικών εικόνων έγινε με την χρήση οπτικού μικροσκοπίου και ενσωματωμένης κάμερας.

Η δεύτερη διαδικασία αφορά την επεξεργασία των ιστοπαθολογικών εικόνων μικροσκοπίας, όπου με τεχνικές τμηματοποίησης απομονώνεται η περιοχή ενδιαφέροντος, οι κυτταρικοί πυρήνες, οι οποίοι φέρουν καφέ χρώμα. Με συνδυασμό τεχνικών, το σύστημα είναι ικανό να προβλέψει και να διαχωρίσει περιοχές ενδιαφέροντος που κρύβουν δύο οι περισσότερους προσκολλημένους κυτταρικούς πυρήνες. Η τρίτη διαδικασία περιλαμβάνει την εξαγωγή των χαρακτηριστικών από την περιοχή ενδιαφέροντος της ιστοπαθολογικής εικόνας. Εξάγουμε 24 χαρακτηριστικά, εκ των οποίων τα 18 είναι χαρακτηριστικά υφής και τα υπόλοιπα 6 μορφολογικά χαρακτηριστικά. Για την εξαγωγή των χαρακτηριστικών υπολογίστηκαν μήτρες όπως, «co-occurrence matrix» και «gray level run length matrix».

Η τέταρτη διαδικασία αφορά την υλοποίηση του συστήματος αναγνώρισης προτύπων. Για την επιλογή χαρακτηριστικών χρησιμοποιήθηκαν τεχνικές όπως η εξαντλητική αναζήτηση βέλτιστων χαρακτηριστικών «exhaustive search» και τεχνικές μη βέλτιστης επιλογής χαρακτηριστικών με στατιστικούς ελέγχους. Για την ταξινόμηση των προτύπων, μελετήθηκαν αλγόριθμοι ταξινόμησης που ήταν ικανοί να διαχωρίσουν άγνωστα πρότυπα μεταξύ τριών διαθέσιμων κλάσεων ( επιπέδων «grades I, II, ή III» ενδομητριακού καρκίνου) όπως ο «Minimum Distance Classifier (MDC)», ο «k-Nearest Neighbours (K-NN)», ο «Least Squares Minimum Distance Classifier (LSMDC)», ο «Linear Bayes Classifier (LBC)» και τα «Probabilistic Neural Network (PNN)». Εκτός από την επιλογή του καλύτερου ταξινομητή, προτάθηκαν τρόποι υλοποίησης με συνδυασμό ταξινομητών «ensemble classifier» κάνοντας χρήση συνδυαστικών τεχνικών και κανόνων απόφασης «majority rules».

Η τελευταία διαδικασία αφορά την αξιολόγηση των τεχνικών και αλγορίθμων που προτάθηκαν. Συγκρίνοντας τα αποτελέσματα, επιλέγεται η καλύτερη τεχνική με την οποία προτείνεται η υλοποίηση του συστήματος. Για τον εγγύτερο υπολογισμό των παραμέτρων εκτίμησης χρησιμοποιήθηκαν μέθοδοι επαναληπτικής δειγματοληψίας, όπως τις τεχνικές «leave one out» και «external cross validation».

Η ανάπτυξη του συστήματος λήψης, επεξεργασίας και ανάλυσης ιστοπαθολογικών εικόνων μικροσκοπίας που περιγράφουμε είναι εφικτή, με πολύ ενθαρρυντικά αποτελέσματα. Με βάση τα αποτελέσματα καλύτερος αλγόριθμος είναι τα πιθανοτικά νευρωνικά δίκτυα. Στην επιλογή βέλτιστου συνδυασμού χαρακτηριστικών με εξαντλητική αναζήτηση, κάνοντας χρήση της μεθόδου «leave one out», επιτεύχθηκε το ποσοστό 96% ορθής ταξινόμησης. Αντίστοιχα είχαμε 94% ορθής ταξινόμησης, στις μη βέλτιστες μεθόδους επιλογής καλύτερου συνδυασμού χαρακτηριστικών. Τέλος στην τελική αξιολόγηση του ταξινομητή επιτεύχθηκε το ποσοστό 82.5%  $\pm$  6.2 ορθής ταξινόμησης άγνωστων προτύπων, με την επαναδειγματοληπτική μέθοδο «external cross validation», επιλέγοντας χαρακτηριστικά με την μη βέλτιστη τεχνική «sequentialfs».

**ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ:** Επεξεργασία εικόνας, Αναγνώριση προτύπων  
**ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ:** Επεξεργασία εικόνας, Τμηματοποίηση εικόνας,  
Χαρακτηριστικά υφής, Μορφολογικά χαρακτηριστικά, Επιλογή χαρακτηριστικών,  
Αναγνώριση Προτύπων, Ενδομητριακός καρκίνος, Ογκογονίδιο *cerb-B*, Οπτικό  
μικροσκόπιο, Ιστοπαθολογικές εικόνες

## ABSTRACT

The objective of the foreclosed thesis is the study and development of a system which receives, processes and analyses histopathological microscopical images, to assist/support the diagnosis of endometrial cancer. From the current available sources, no other study was found with regard to the quantitative analysis of histopathological images related to endometrial cancer. Similar studies regarding complete systems of histopathological image classification of another type of cancer. The structure of implementation is divided into five basic procedures.

The first one concerns the digitation of available biopsies of endometrial cancer. The biopsies were processed by histopathologist staining Hematoxylin-losinis to assess the degree of tumor and immunohistochemically for the expression of the oncogene *cerb-B*. We have digitized 17 biopsy / incidents involving different patients, from which 117 digital histopathologic microscopical images were taken. The histopathological image sample was taken using an optical microscope and an integrated camera.

The second procedure concerns the processing of histopathological microscopical images by segmenting the region of interest (ROI). The ROI is defined as the expressed nuclei that where colored brown. With a combination of techniques, the system is able to predict and distinguish ROIs that conceal two or more adherent nuclei.

The third procedure involves the feature extraction stage. We extract 24 features, of which 18 are textural features and the other 6 are morphological features. Textural features were calculated based on the gray level co-occurrence matrix and the gray level run length matrix.

The fourth procedure concerns the implementation of the pattern recognition system. For the features selection stage, the Kruskal-Wallis statistical test and a linear regression method were employed as non optimal techniques. The exhaustive search method was also utilized as the optimum feature selection technique. For the classification stage, a variety of classifiers were tested such as the Minimum Distance (MDC), the k-Nearest Neighbour (K-NN), the Least Squares Minimum Distance (LSMDC), the 'Linear Bayes (LBC) and the Probabilistic Neural Network (PNN). The classification task was to characterize an image into three grades of endometrial cancer (grade I, II, or III). Apart from the classical implementation of the pattern recognition system of choosing the best classifier, methods of implementation combining classifiers 'ensemble classification schemes' were proposed, using combinatorial techniques and decision rules 'majority rules'.

The last procedure concerns the evaluation of the techniques and algorithms which have been proposed. Comparing the results, the best technique which the system implementation proposes, is selected. For the best estimation of the calculating parameters, methods of iterative sampling techniques were used, such as 'leave one out' and 'external cross validation'.

Best classification algorithm was found the PNN that gave 94% of correct classification when non optimal feature selection method were employed. The PNN scored around 96% overall accuracy employing the 'exhaustive search' and the 'leave one out' methods. The proposed image analysis system proved

capable of classifying a 'new' image with an average accuracy of 82% employing the 'external cross validation' method. Results are promising for the development of such system, with some modest modifications suitable for a clinical environment.

**SUBJECT AREA:** Digital image processing , Pattern Recognition

**KEYWORDS:** Image Processing, Image Segmentation, Texture features, Morphological features , Features selection, Pattern Recognition, Endometrial cancer, Oncogene *cerb-B*, Optical microscope, Histopathological image,