



**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 καρκίνου θυρεοειδούς αδένα  
από κυτταρολογικές εικόνες**

**Βασιλική Γ. Τσιμπίδα**

**Επιβλέπων: Διονύσης Κάβουρας, Καθηγητής ΤΕΙ-Α**

**ΑΘΗΝΑ**

**ΙΟΥΛΙΟΣ 2013**

## **ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

Σύστημα αυτόματης διάγνωσης καρκίνου θυρεοειδούς αδένα από κυτταρολογικές  
εικόνες

**Βασιλική Γ. Τσιμπίδα**

**A.M.: ΠΙΒ11 - 069**

**ΕΠΙΒΛΕΠΩΝ:** Διονύσης Κάβουρας, Καθηγητής ΤΕΙ-Α

**ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ:** Διονύσης Κάβουρας, Καθηγητής ΤΕΙ-Α  
Μανώλης Σαγκριώτης, Αναπληρωτής Καθηγητής ΕΚΠΑ  
Σπύρος Κωστόπουλος, Μεταδιδακτορικός ερευνητής ΤΕΙ-Α

Ιούλιος 2013

## ΠΕΡΙΛΗΨΗ

Αντικείμενο της παρούσας εργασίας είναι η ανάπτυξη αλγορίθμων αναγνώρισης προτύπων για το σχεδιασμό συστήματος ταξινόμησης καρκίνου του θυρεοειδούς αδένου (καλοήθεια-κακοήθεια), μέσω της επεξεργασίας και ανάλυσης κυτταρολογικών εικόνων.

Η διαδικασία περιλαμβάνει την ψηφιοποίηση εικόνων μικροσκοπίας από πρωτογενές υλικό, σε ειδικά παρασκευασμένα πλακίδια, επεξεργασμένο με χρώση Αιματοξυλίνης-Ιωσίνης. Το υλικό περιλαμβάνει είκοσι δείγματα βιοψίας θυρεοειδούς αδένου από αντίστοιχα περιστατικά, διαγνωσμένα από έμπειρο ιστοπαθολόγο ιατρό.

Για το σχεδιασμό και την υλοποίηση του συστήματος χρησιμοποιήθηκαν δέκα (10) εικόνες καλοήθειας και δέκα (10) εικόνες κακοήθειας θυρεοειδούς αδένου. Το σύστημα αρχικά περιλαμβάνει αλγόριθμους προ-επεξεργασίας και τμηματοποίησης κυτταρολογικών εικόνων για την εύρεση των περιοχών ενδιαφέροντος (πυρήνες). Συγκεκριμένα, μελετήθηκαν αλγόριθμοι τμηματοποίησης εικόνων μικροσκοπίας (κατωφλίωση με σταθερό κατώφλι, κατωφλίωση με προσαρμοζόμενο κατώφλι, μέθοδος Otsu) για την εύρεση της βέλτιστης λύσης και συγκρίθηκαν με αντίστοιχα εμπορικά πακέτα, όπως ImPro, όπου παρατηρήθηκε ότι οι εικόνες αυτού του πακέτου ήταν ευδιάκριτες σε πολύ μεγάλο βαθμό, χωρίς αλληλοεπικαλύψεις κυτταρικών πυρήνων. Χρησιμοποιώντας τα αποτελέσματα του ImPro, από τις εικόνες καλοήθειας προέκυψαν 217 τμηματοποιημένοι κυτταρικοί πυρήνες και από τις εικόνες κακοήθειας 328 τμηματοποιημένοι κυτταρικοί πυρήνες. Στη συνέχεια από αυτούς τους πυρήνες έγινε εξαγωγή πέντε (5) μορφολογικών χαρακτηριστικών και δώδεκα (12) χαρακτηριστικών υψής. Κατόπιν, υλοποιήθηκαν οι ταξινομητές Ελάχιστης Απόστασης, Πλησιέστερου Γείτονα, Μπαεσσανός και Πιθανοκρατικό Νευρωνικό Δίκτυο για την ταξινόμηση των κυτταρικών πυρήνων σε δύο κατηγορίες. Για κάθε έναν ταξινομητή, βρέθηκε ο βέλτιστος συνδυασμός χαρακτηριστικών, εφαρμόζοντας μέθοδο επιλογής χαρακτηριστικών τη Sequential Backward Selection και την εξαντλητική αναζήτηση. Το προτεινόμενο σύστημα αξιολογήθηκε με Leave-One-Out μέθοδο αξιολόγησης. Η ακρίβεια του συστήματος σε 'νέα' δεδομένα αξιολογήθηκε με την External Cross Validation.

Οχτώ (8) χαρακτηριστικά παρουσίασαν στατιστικά σημαντικές διαφορές ( $p < 0.001$ ) σύμφωνα με το στατιστικό τεστ Wilcoxon. Το σύστημα ταξινόμησε με 95% επιτυχία τους κυτταρικούς πυρήνες στις δύο κατηγορίες, χρησιμοποιώντας το ταξινομητή 3 Πλησιέστερων Γειτόνων. Το προτεινόμενο σύστημα ταξινομεί 'νέους' κυτταρικούς πυρήνες με 93% ακρίβεια.

**ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ:** Επεξεργασία και Ανάλυση Εικόνας

**ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ:** τμηματοποίηση, εξαγωγή χαρακτηριστικών, αναγνώριση προτύπων, καρκίνος θυρεοειδούς, μικροσκοπία

## 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 unknown' nuclei with 93% accuracy.

**SUBJECT AREA:** Image Processing and Analysis

**KEYWORDS:** segmentation, feature extraction, pattern recognition, thyroid cancer, microscopy