



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 προ-καρκινικών
αλλοιώσεων λόγω HPV-λοίμωξης τραχήλου μήτρας από
ιστοπαθολογικές εικόνες**

Αικατερίνη Β. Ρώτα

Επιβλέπων: Διονύσης Κάβουρας, Καθηγητής ΤΕΙ-Α

ΑΘΗΝΑ

ΑΠΡΙΛΙΟΣ 2014

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Σύστημα αυτόματης διάγνωσης προ-καρκινικών αλλοιώσεων λόγω HPV-λοίμωξης
τραχήλου μήτρας από ιστοπαθολογικές εικόνες

Αικατερίνη Β. Ρώτα
Α.Μ.: ΠΙΒ11077

ΕΠΙΒΛΕΠΩΝ: Διονύσης Κάβουρας, Καθηγητής ΤΕΙ-Α

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

Απρίλιος 2014

ΠΕΡΙΛΗΨΗ

Αντικείμενο της παρούσας διπλωματικής εργασίας είναι ο σχεδιασμός ενός συστήματος αυτόματης διάγνωσης για την ταξινόμηση των προ-καρκινικών αλλοιώσεων του τραχήλου της μήτρας σε χαμηλού (CIN I) και υψηλού βαθμού επικινδυνότητας (CIN II, CIN III), με την επεξεργασία ιστολογικών εικόνων, ώστε να βελτιστοποιηθεί η θεραπευτική αντιμετώπιση των ασθενών.

Συγκεκριμένα, χρησιμοποιήθηκαν τριάντα πέντε (35) δείγματα ασθενών (βιοψίες), που διαγνώστηκαν με ενδο-επιθηλιακές αλλοιώσεις. Από τα συνολικά δείγματα, τα δώδεκα (12) διαγνώστηκαν ως χαμηλού βαθμού επικινδυνότητα (CIN1), ενώ τα υπόλοιπα είκοσι τρία (23) ως υψηλού βαθμού επικινδυνότητα (CIN2 και CIN3), σύμφωνα με τη γνωμάτευση έμπειρου ιστοπαθολόγου, που χρησιμοποίησε το σύστημα ταξινόμησης του Παγκόσμιου Οργανισμού Υγείας.

Αρχικά, πραγματοποιήθηκε η προετοιμασία των δεδομένων, κατά την οποία τα δείγματα ιστού χρωματίστηκαν με H&E (Hematoxylin & Eosin) και ακολούθησε η ψηφιοποίηση της εικόνας, όπου από κάθε δείγμα ψηφιοποιήθηκαν εικόνες, οι οποίες πάρθηκαν από την αντιπροσωπευτική περιοχή (Region of Interest (ROI)), που όρισε ο ιστοπαθολόγος. Ακολούθησε τμηματοποίηση των πυρήνων της εικόνας με τη μέθοδο της κατωφλίωσης και τη χρήση του τελεστή μορφολογίας opening. Κατόπιν, από τους τμηματοποιημένους πυρήνες έγινε εξαγωγή δεκαεπτά (17) χαρακτηριστικών, πέντε (5) μορφολογικών και δώδεκα (12) χαρακτηριστικών υψής.

Τα χαρακτηριστικά αυτά αποτέλεσαν την είσοδο σε ένα σύστημα αναγνώρισης προτύπων που σχεδιάστηκε, έτσι ώστε να προβλέψει τον τύπο ενδοεπιθηλιακής αλλοίωσης. Δοκιμάστηκαν οι ταξινομητές: «k-Nearest Neighbors (k-NN)», «Probabilistic Neural Network (PNN)», «Minimum Distance Classifier (MDC)», «Least Squares Minimum Distance Classifier (LSMDC)», «Support Vector Machines (SVM)» και «Single Perceptron» για την ταξινόμηση των εικόνων σε δύο κατηγορίες (χαμηλού ή υψηλού βαθμού επικινδυνότητα).

Για καθέναν ταξινομητή, βρέθηκε ο βέλτιστος συνδυασμός χαρακτηριστικών, εφαρμόζοντας ως μεθόδους επιλογής χαρακτηριστικών, τη Sequential Backward Selection και την Exhaustive Search. Το σύστημα αξιολογήθηκε με τη μέθοδο αξιολόγησης Leave-One-Out (LOO). Η γενίκευση του συστήματος σε “άγνωστα” δεδομένα αξιολογήθηκε με τη μέθοδο External Cross Validation.

Τα χαρακτηριστικά λοξότητα, κυρτότητα, αντίθεση, συσχέτιση, έμφαση στις μικρές δομές, η ανομοιομορφία τόνων του γκρι και ανομοιομορφία μήκους διαδρομής παρουσίασαν στατιστικά σημαντικές διαφορές ($p < 0.001$) μεταξύ των δυο κατηγοριών, χρησιμοποιώντας το Wilcoxon μη παραμετρικό στατιστικό τεστ.

Το προτεινόμενο σύστημα ανάλυσης εικόνας και αναγνώρισης προτύπων φαίνεται να διαχωρίζει εικόνες προ-καρκινικών αλλοιώσεων του τραχήλου της μήτρας σε χαμηλού (CIN I) και υψηλού βαθμού επικινδυνότητας (CIN II, CIN III) με 94,6-95% ακρίβεια χρησιμοποιώντας τον ταξινομητή PNN, και την ECV μέθοδο αξιολόγησης. Τα χαρακτηριστικά που έχουν μεγάλη συχνότητα εμφάνισης στον PNN ταξινομητή είναι το εμβαδό, η κύρτωση, η εκκεντρικότητα και η ανομοιομορφία γκρι επιπέδου. Αυτά τα χαρακτηριστικά περιγράφουν ομοιογένεια και μορφολογία στην εικόνα

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Επεξεργασία εικόνας, Αναγνώριση Προτύπων

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: τράχηλος μήτρας, ενδοεπιθηλιακή αλλοίωση (CIN I, II, III), κυτταρολογική εικόνα, τμηματοποίηση εικόνας, μορφολογικά χαρακτηριστικά, χαρακτηριστικά υφής, επιλογή χαρακτηριστικών, αναγνώριση προτύπων

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: Image Processing, Pattern Recognition

KEYWORDS: cervix, intraepithelial lesion (CINI, II, III), histopathological image, image segmentation, morphological characteristics, textural characteristics, feature selection, pattern recognition