

**ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

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

Ρώτα Αικατερίνη

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

Περίληψη

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 υφής.

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