



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

**Μέθοδοι υπολογιστικής νοημοσύνης για την ταξινόμηση
μαζών από μαστογραφικές εικόνες**

Πάυλος Π. Καφούρης

Επιβλέπων: Γεώργιος Σπύρου, Ειδικός Λειτουργικός Επιστήμονας (βαθμίδα Α')

ΑΘΗΝΑ

ΟΚΤΩΒΡΙΟΣ 2014

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Μέθοδοι υπολογιστικής νοημοσύνης για την ταξινόμηση μαζών από μαστογραφικές εικόνες

Παύλος Π. Καφούρης

A.M.: ΠΙΒ 090

ΕΠΙΒΛΕΠΩΝ: Γεώργιος Σπύρου, Ειδικός Λειτουργικός Επιστήμονας (βαθμίδα Α')

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ: Γεώργιος Σπύρου, Ειδικός Λειτουργικός Επιστήμονας
Εμμανουήλ Σαγκριώτης, Αναπληρωτής Καθηγητής
Εμμανουήλ Αθανασιάδης, Μεταδιδακτορικός Ερευνητής

Οκτώβριος 2014

ΠΕΡΙΛΗΨΗ

Ο όρος «καρκίνος του μαστού» αναφέρεται στην ανάπτυξη κακοήθους όγκου στην περιοχή του μαστού. Αποτελεί μία από τις συχνότερα εμφανιζόμενες μορφές καρκίνου παγκοσμίως και είναι η πρώτη σε αριθμό κρουσμάτων στο γυναικείο πληθυσμό. Η πιθανότητα εμφάνισης της νόσου σε άνδρες είναι υπαρκτή, αλλά πολύ μικρή. Το πιο αξιόπιστο εργαλείο για την ανίχνευση και τη διάγνωση του καρκίνου του μαστού αποτελεί η μαστογραφία.

Στην παρούσα διπλωματική εργασία στόχος είναι η ανάπτυξη μεθοδολογιών που μπορούν να χρησιμοποιηθούν σε ένα αυτόματο σύστημα υποβοηθούμενης διάγνωσης (CADx) για μάζες που εντοπίζονται σε μαστογραφίες. Η μάζα είναι ένα σύμπλεγμα κυττάρων με μεγαλύτερη πυκνότητα από τον περιβάλλοντα ιστό, με αποτέλεσμα να είναι σχετικά εύκολα παρατηρήσιμη στις μαστογραφίες. Η υλοποίηση του συστήματος περιλαμβάνει μία σειρά αυτόνομων σταδίων όπως: η προεπεξεργασία της εικόνας στοχεύοντας στη βελτίωση της ποιότητάς της και στον εντοπισμό σημαντικών ανατομικών δομών του μαστού, η κατάτμηση των μαζών από την αρχική εικόνα, η εξαγωγή χαρακτηριστικών περιγραφής τους, στη συνέχεια η επιλογή των πιο σημαντικών από αυτά και τέλος η εφαρμογή μεθόδων ταξινόμησης, ώστε να επιτευχθεί με αυτόματο τρόπο ο διαχωρισμός μεταξύ των καλοήθων και κακοήθων μαζών. Μετά την ανάπτυξη των αυτόνομων σταδίων του συστήματος, ακολουθεί η αξιολόγησή του με σκοπό τη διερεύνηση της αποτελεσματικότητάς του και τη δυνατότητα εφαρμογής του σε συνθήκες κλινικής πρακτικής. Για την αξιολόγηση του συστήματος χρησιμοποιήθηκαν 901 εικόνες από πραγματικές περιπτώσεις ασθενών που ελήφθησαν από την ελεύθερα διαθέσιμη βάση δεδομένων DDSM, εξασφαλίζοντας με τον τρόπο αυτό αντικειμενικά αποτελέσματα, αφού πρόκειται για μεγάλο πλήθος περιπτώσεων. Τα μέτρα που χρησιμοποιήθηκαν για τον έλεγχο της επίδοσης των ταξινομητών σε τελείως άγνωστα δεδομένα (270 περιπτώσεις) είναι η ακρίβεια, η ευαισθησία, η ειδικότητα του κάθε ταξινομητή, καθώς και το εμβαδόν κάτω από την καμπύλη ROC.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Επεξεργασία Μαστογραφικών Εικόνων

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: μαστογραφία, σύστημα υποβοηθούμενης διάγνωσης, μάζα, κατάτμηση μαστού, ταξινόμηση

ABSTRACT

The term “breast cancer” refers to a malignant tumor developing from breast tissue. Worldwide it is among the most common cancers and its incidences come first among the women population. The probability that breast cancer occurs in males exists, but it is quite small. Mammography is the most reliable technique for screening and diagnosing breast cancer.

The aim of this master’s thesis is the development of methodologies that may be integrated in a computer aided diagnosis system (CADx) for masses detected in mammographic images. Mass consists of a complex of cells, easily distinguishable through mammography, due to its higher than the surrounding tissue density. The development of the system includes various independent steps such as: image preprocessing, aiming to the enhancement of image quality and the detection of important mammographic structures, the segmentation of the masses from the original image, the extraction of their features, then the application of feature selection and finally the application of classification methods for the discrimination between benign and malignant masses. After developing all subunits of the system, we proceed on its evaluation in order to investigate its efficacy and its potential to be adopted in daily clinical practice. For the evaluation of the system were used 901 images, from real-life cases taken from the public DDSM database, achieving consequently objective results, since it refers to a satisfying number of cases. In order to evaluate the classifiers’ performance at totally unknown data (270 cases) the factors that were used are accuracy, sensitivity, specificity of each classifier, as well as calculation of the area under the ROC curve.

SUBJECT AREA: Mammographic Image Processing

KEYWORDS: mammography, computer aided diagnosis system, mass, breast segmentation, classification