



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 πνευμονικών όζων
από εικόνες αξονικής τομογραφίας**

Πέγκυ Β. Λάγγα

Επιβλέπουσα: Γεωργία Οικονόμου, Αναπληρώτρια Καθηγήτρια

ΑΘΗΝΑ

ΑΠΡΙΛΙΟΣ 2012

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Σύστημα αυτόματης διάγνωσης πνευμονικών όζων από εικόνες αξονικής τομογραφίας

Πέγκυ Β. Λάγγα

A.M.: ΠΙΒ037

ΕΠΙΒΛΕΠΟΥΣΑ: Γεωργία Οικονόμου, Αναπληρώτρια Καθηγήτρια

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ: Διονύσης Κάβουρας, Καθηγητής
Εμμανουήλ Σαγκριώτης, Αναπληρωτής Καθηγητής
Γεωργία Οικονόμου, Αναπληρώτρια Καθηγήτρια

Απρίλιος 2012

ΠΕΡΙΛΗΨΗ

Η διπλωματική αυτή εργασία έχει σκοπό την υλοποίηση ενός συστήματος αυτόματης διάγνωσης όζων του πνεύμονα, ώστε να παραχθεί ένα μοντέλο το οποίο θα λειτουργεί συμπληρωματικά στην ιατρική γνωμάτευση, παρέχοντας όσο το δυνατό ακριβέστερη διάγνωση για την ανίχνευση του καρκίνου του πνεύμονα. Η μέθοδος που εφαρμόζεται στηρίζεται στην επεξεργασία και την ανάλυση των εικόνων της αξονικής τομογραφίας πνεύμονα, επιδιώκοντας την ταξινόμησή τους.

Αναλυτικότερα, χρησιμοποιήθηκαν εικόνες αξονικής τομογραφίας πνεύμονα από 12 ασθενείς, γνωματευμένες από έμπειρο ακτινολόγο γιατρό. Οι εικόνες αυτές προεπεξεργάστηκαν, ώστε να απομονωθούν οι πνεύμονες από τον περιβάλλοντα ιστό. Έπειτα, προκειμένου να γίνει η ταξινόμηση των περιοχών του πνεύμονα, ακολουθήσαμε διάφορα βήματα, όπως επιλογή των περιοχών ενδιαφέροντος (ROIs), εξαγωγή χαρακτηριστικών υφής, μείωση και επιλογή χαρακτηριστικών, ταξινόμηση. Αφού λοιπόν έγινε εξαγωγή 20 χαρακτηριστικών υφής από τις περιοχές ενδιαφέροντος και κατόπιν μείωσή τους, ώστε να κερδίσουμε χρόνο και υπολογιστικό κόστος, χωρίσαμε τα συνολικά δεδομένα μας (υγιή και καρκινικά) σε δύο ομάδες, την εκπαιδευτική και τη δοκιμαστική. Τα δεδομένα της εκπαιδευτικής ομάδας χρησιμοποιήθηκαν για το σχεδιασμό του ταξινομητή, με τον οποίο έγινε ο χαρακτηρισμός των δεδομένων της δοκιμαστικής ομάδας.

Όσον αφορά την ταξινόμηση των δεδομένων σε καρκινικά και υγιή, υλοποιήσαμε ένα σύστημα αναγνώρισης προτύπων με 7 διαφορετικούς αλγόριθμους ταξινόμησης (MDC, KNN unweighted, KNN weighted, PNN Gaussian, PNN exponential, PNN reciprocal, SVM). Εφαρμόζοντας για κάθε έναν ταξινομητή εξαντλητική αναζήτηση και κάνοντας χρήση της μεθόδου leave one out, καταλήξαμε στον καλύτερο συνδυασμό χαρακτηριστικών, από τον οποίο προέκυπτε το μεγαλύτερο ποσοστό ακρίβειας. Επιπλέον, υπολογίστηκε το ποσοστό της ακρίβειας ταξινόμησης για το συνδυασμό και των 7 ταξινομητών ίσο με 90%, κάνοντας χρήση του κανόνα της πλειοψηφίας.

Τέλος, αφού συγκρίναμε τα αποτελέσματα των ταξινομητών, καταλήξαμε σε μία ομάδα 5 ταξινομητών (MDC, KNN unweighted, KNN weighted, PNN exponential, SVM), ο συνδυασμός των οποίων ενισχύει την απόδοση του συστήματος. Σχεδιάσαμε λοιπόν ξανά τους ταξινομητές μας, με το σύνολο των δεδομένων μας αυτή τη φορά, αλλά χρησιμοποιώντας μόνο τους βέλτιστους συνδυασμούς χαρακτηριστικών τους, και ταξινομήσαμε κάθε σημείο της εικόνας της αξονικής τομογραφίας κάνοντας χρήση του κανόνα της πλειοψηφίας για τους 5 καλύτερους ταξινομητές.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Επεξεργασία και ανάλυση ιατρικής εικόνας

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: Καρκίνος του πνεύμονα, ψηφιακή επεξεργασία ιατρικής εικόνας, επιλογή χαρακτηριστικών υφής, ταξινόμηση προτύπων, σύστημα υποβοήθησης λήψης αποφάσεων

ABSTRACT

This thesis aims to implement a computerized system for the detection of lung nodules, in order to provide an accurate diagnosis support model for lung cancer screening. This method is based on processing and analyzing images of computed tomography (CT) of the lungs, so as to classify these images.

Specifically, there were used CT images of lungs from 12 patients, which were assessed by an experienced radiologist. These images were digitized and were segmented in order to distinguish lungs from background tissue. Then, in order to classify the areas of the lung, we followed various steps, such as selection of regions of interest (ROIs), extraction of textural features, reduction and selection of features, classification. After the extraction of 20 textural features from the regions of interest, we reduced them, so as to save time and computational cost, and we divided the whole data (healthy and nodules) in two groups, training and test. The training data were used to design the classifier, with whom we did the characterization of the test data.

With regard to the classification of the data as cancer or healthy, we created a pattern recognition system where 7 different classifiers were tested (MDC, KNN unweighted, KNN weighted, PNN Gaussian, PNN exponential, PNN reciprocal, SVM). We then applied the exhaustive search method in each classifier and, also, by using the leave one out method, we resulted to the best feature combination, which gives the highest percentage of accuracy. Moreover, it was calculated the percentage of accuracy equal to 90%, for the classification which was occurred by the combination of the 7 classifiers, using the majority rule.

Finally, having compared the results of accuracy for the classifiers, we ended up with a group of 5 classifiers (MDC, KNN unweighted, KNN weighted, PNN exponential, SVM), a combination that enhanced system's performance. So, we designed our classifiers once again, with the whole number of our data this time and using only the best feature combinations, and we classified every pixel of the CT of lungs, using the majority rule for the best 5 classifiers.

SUBJECT AREA: Processing and analysis of medical image

KEYWORDS: Lung cancer, digital processing of medical image, selection of textural features, pattern classification, decision support system