



**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**

**Ανάπτυξη ολοκληρωμένου περιβάλλοντος ανάλυσης  
και ταξινόμησης μαστογραφικών εικόνων**

**Κωνσταντίνα Χ. Λεβέντη**

**Επιβλέποντες: Γεώργιος Σπύρου, Ειδικός Λειτουργικός Επιστήμονας Α'**

**ΑΘΗΝΑ**

**ΔΕΚΕΜΒΡΙΟΣ 2014**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

**Ανάπτυξη ολοκληρωμένου περιβάλλοντος ανάλυσης  
και ταξινόμησης μαστογραφικών εικόνων**

**Κωνσταντίνα Χ. Λεβέντη**

**A.M.: ΠΙΒ 091**

**ΕΠΙΒΛΕΠΟΝΤΕΣ:** Γεώργιος Σπύρου, Ειδικός Λειτουργικός Επιστήμονας Α'

**ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ:** Γεωργία Οικονόμου, Αναπληρώτρια Καθηγήτρια  
Εμμανουήλ Αθανασιάδης, Μεταδιδακτορικός Ερευνητής

Δεκέμβριος 2014

## ΠΕΡΙΛΗΨΗ

Ο καρκίνος του μαστού είναι μια από τις συχνότερες μορφές καρκίνου στον γυναικείο πληθυσμό. Η μαστογραφία είναι μια αποτελεσματική και ασφαλής μέθοδος για την ανίχνευση και διάγνωση αυτής της ασθένειας. Ωστόσο, η ερμηνεία των μαστογραφιών ενέχει συχνά δυσκολίες για τους ακτινολόγους που αφορούν παράγοντες όπως η κακή ποιότητα της εικόνας ή η απροσεξία των ίδιων των γιατρών. Έτσι, τις τελευταίες δεκαετίες έχουν αναπτυχθεί συστήματα υποβοηθούμενης διάγνωσης (CAD) που παρέχουν στους ακτινολόγους μια δεύτερη γνώμη για να διαμορφώσουν την τελική τους διάγνωση.

Ένα από τα σημαντικότερα ευρήματα που σχετίζεται με τον καρκίνο του μαστού είναι οι μικροασβεστώσεις (MCs), μικρές δομές που δημιουργούνται από συγκέντρωση αλάτων ασβεστίου και εμφανίζονται συνήθως με την μορφή συμπλεγμάτων. Τα ευρήματα αυτά μπορεί να είναι τόσο καλοήθη όσο και κακοήθη και έχουν γίνει πολλές προσπάθειες για να ταξινομηθούν ανάλογα με τα χαρακτηριστικά τους.

Στην παρούσα διπλωματική εργασία παρουσιάζουμε συνολικά τη λειτουργία του συστήματος υποβοηθούμενης διάγνωσης με την ονομασία Ιπποκράτης-μστ που αναπτύχθηκε στο Εργαστήριο Πληροφορικής της Ακαδημίας Αθηνών. Το σύστημα στηρίζεται στην ανάλυση και αξιολόγηση μεμονωμένων μικροασβεστώσεων, καθώς και αντίστοιχων συμπλεγμάτων. Η υλοποίηση γίνεται σε τέσσερα στάδια: α) αρχειοθέτηση ασθενών, β) χρήση τεχνικών ανάλυσης εικόνας για την εξέταση των μαστογραφιών, γ) ανίχνευση και ανάλυση των μικροασβεστώσεων και δ) εξαγωγή διάγνωσης (καλοήθεια ή κακοήθεια). Η φάση της διάγνωσης στηρίζεται στον αλγόριθμο μηχανικής μάθησης των support vector machines (SVM). Η μεθοδολογία SVM επιλέχθηκε καθώς έχει καλύτερα αποτελέσματα από άλλους αλγορίθμους μηχανικής μάθησης όταν εφαρμόζεται σε δεδομένα που δεν ανήκουν στο σύνολο εκπαίδευσης, συνεπώς έχει μεγαλύτερη ικανότητα γενίκευσης και βρίσκει ταχύτερα την βέλτιστη λύση.

Στα πλαίσια της εργασίας αυτής αναπτύχθηκε ένα συνδυαστικό σχήμα ταξινόμησης μαστογραφικών εικόνων ως καλοήθεις ή κακοήθεις, το οποίο αποτελείται από έναν SVM ταξινομητή και έναν νέο ταξινομητή που δημιουργήσαμε. Ο SVM εκπαιδεύεται με ένα μικρό σύνολο χαρακτηριστικών των μικροασβεστώσεων που επελέγησαν μετά από υπολογισμούς. Ο νέος ταξινομητής προσφέρει δυνατότητα κατηγοριοποίησης νέων μαστογραφικών εικόνων με βάση το περιεχόμενό τους και στηρίζεται στον υπολογισμό αποστάσεων ανάμεσα στο διάνυσμα χαρακτηριστικών της άγνωστης εικόνας και των γνωστών εικόνων. Η απόφαση προκύπτει από τις ψήφους των κοντινότερων γνωστών εικόνων. Η τελική πρόβλεψη της άγνωστης εικόνας,

προκύπτει από τον συνδυασμό των προβλέψεων των δυο ταξινομητών, με εφαρμογή ενός απλού κανόνα.

Επίσης, επικαιροποιήθηκε η βάση MIRACLE ώστε να είναι δυνατή η διασύνδεση με το μετασχηματισμένο σύστημα του Ιπποκράτη-μστ που λειτουργεί σε περιβάλλον Java για να είναι προσιτό ως προς τη χρήση του.

Στο τέλος της εργασίας, παρουσιάζεται μια ολοκληρωμένη λύση για την υποβοήθηση της μαστογραφικής διάγνωσης, μαζί με την αξιολόγησή της, καθώς και τα συμπεράσματα που προέκυψαν από όλη αυτή τη μελέτη.

**ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ:** Επεξεργασία Εικόνας, Ανάλυση Εικόνας, Ταξινόμηση μαστογραφικής εικόνας, Υπολογιστική Υποβοηθούμενη Διάγνωση

**ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ:** Ταξινόμηση, Απεικονιστικά Ευρήματα, Αλγόριθμος SVM, Content-based Αλγόριθμος, DDSM Database, Καλοήθεια, Κακοήθεια, Ιπποκράτης-μστ, Σύστημα CAD

## ABSTRACT

Breast cancer is one of the most common cancers in the female population. Mammography is an effective and safe method to detect and diagnose this disease. However, the interpretation of mammograms often involves difficulties for the radiologists that depend on factors such as bad quality of the images, or human fatigue. Hence, during the last decades Computer Aided Diagnosis (CAD) systems have been developed that provide radiologists a second opinion to help them form their final diagnosis.

One of the most important findings related to breast cancer is Microcalcifications (MCs), small structures that are created from concentrated calcium salts and usually appear in the form of clusters. These findings can be either benign or malignant and many efforts have been made to classify them according to their features.

In this master thesis we present the functionality of the CAD system named Hippocrates-mst that has been developed at the Informatics Laboratory of the Academy of Athens. The system is based on the analysis and evaluation of both single microcalcifications and clusters of MCs. The implementation is done in four steps: a) patient's archive analysis, b) image analysis techniques for the examination of mammographies, c) detection and analysis of MCs and d) final diagnosis (benign or malignant). The diagnosis phase that we have used is the one based on the machine learning algorithm of support vector machines (SVM).

During this project a combined classification scheme has been developed in order to classify mammographic images as benign or malignant. This scheme consists of an SVM classifier and a new classifier that we have created. The SVM is trained with a small group of MCs features that were selected after numerous calculations. The other classifier provides the ability to categorize new mammograms based on their content and relies on the calculation of distances between the feature vector of the unknown image and the known images. The decision is based on majority voting regarding the nearest known images. The final prediction arises from the combination of the predictions of the two classifiers by applying a simple rule.

In addition, the MIRACLE database has been updated so that it can be connected with the transformed CAD system Hippocrates-mst operating in Java platform in order to become easily accessible.

At the end of this master thesis, we present a full solution for the assistance of mammographic diagnosis, along with its validation and the related conclusions.

**SUBJECT AREA:** Image processing, Image analysis, Mammographic image classification, Computer Aided Diagnosis

**KEYWORDS:** Classification, Image findings, SVM Algorithm, Content-based Algorithm, DDSM Database, Benign, Malignant, Hippokrates-mst, CAD system