Development of an integrated environment of mammographic images analysis and classification

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 Hippoctates-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.

TOPIC: Image processing, Image analysis, Mammographic image classification, Computer Aided Diagnosis
KEY WORDS: Classification, Image findings, SVM Algorithm, Content-based Algorithm, DDSM Database, Benign, Malignant, Hippokrates-mst, CAD system