A hybrid multi-modal computer-aided diagnostic tool for improving diagnostic accuracy in breast cancer classification

Spiros Kostopoulos[1], Dimitris Glotsos[1], Ioannis Kalatzis[1], Pantelis Asvestas[1], Konstantinos Sidiropoulos[2], Panagiota Ravazoula[3], and Dionisis Cavouras[1]

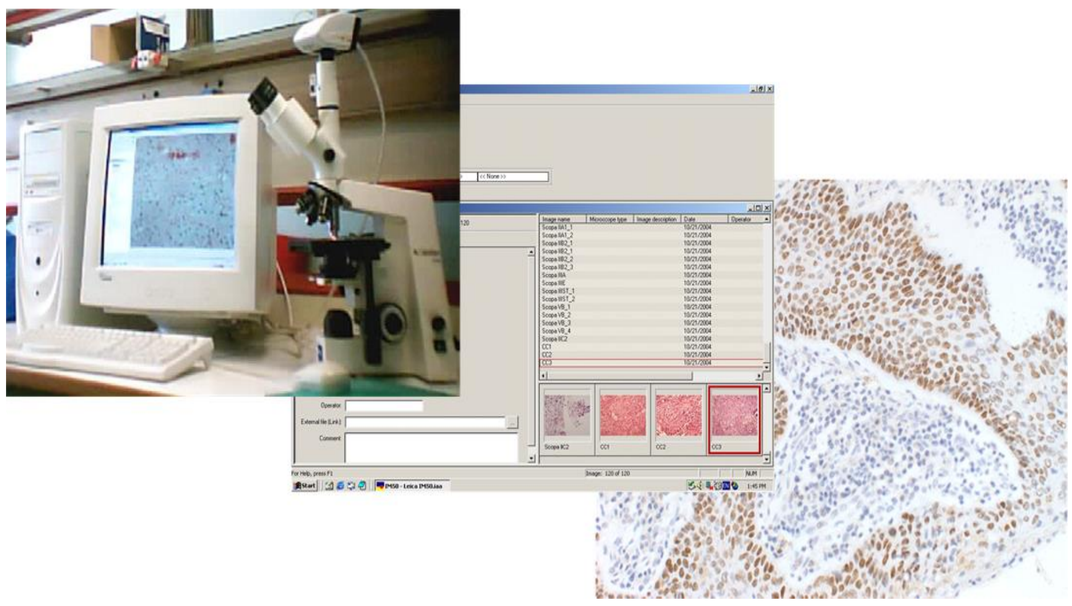
[1]Medical Image and Signal Processing Laboratory, Department of Biomedical Engineering, Technological Educational Institute of Athens, Greece
[2]EMBL-EBI, European Molecular Biology Laboratory, Welcome Trust Genome Campus, Hinxton, Cambridge, UK
[3]Department of Pathology, University Hospital of Patras, Greece

Introduction

The aim of this study is to design, develop and implement in the clinical routine a hybrid multi-modal (radiological and microscopic imaging) computer-aided diagnostic tool for improving diagnostic, prognostic, and predictive classification of patients' breast cancer tumors. This hybrid-multi modal tool was developed under the project CADMAMMO, which has been co-funded by the European Union (European Social Fund) and Greek national resources under the framework of the "Archimedes III: Funding of Research Groups in TEI of Athens" project of the "Education & Lifelong Learning" Operational Programme.

Material

The study considered 71 breast cancer cases of three different grades (I, II or III) collected from the archives of the University Hospital of Patras, Greece. A light microscopy imaging system (LEICA DM 2500 microscope coupled with a LEICA DFC 420C camera, Leica Microsystems GmbH) was used to digitize images from the most representative parts of the tumor, indicated by the expert histopathologist, at 400x and 200x magnifications.



Methods

Feature Extraction: comprised 22 textural (4 from the grey level histogram, 13 from the grey level co-occurrence matrix and 5 from the grey level run length matrix), 11 morphological (Area, Perimeter, Eccentricity, Major Axis Length, Minor Axis Length, Convex Area, Equivalent Diameter, Solidity, Rectangularity, Compactness and Fractal-Dimension), 6 Shape-based on Boundary (Radial Distance (average), Radial Distance (standard deviation), Radial Distance (range), Circularity Ratio, Entropy Radial Distance, Roughness-Index), 7 from the Minimal Spanning Trees (MST) (average Distance-MST, range Distance-MST, standard deviation Distance-MST, max Distance-MST, min Distance-MST, sum Distance-MST, number of Nodes-MST), 6 Molecular Indices (ER, PR, cerb2, p53, Ki-67, cath-D) and 2 mammographic features (shading/vagueness of mCa/lesion, mCa/lesion size in mammography).

Classification: the Least Squares Minimum Distance (LSMD), the k-Nearest Neighbors (kNN), the Probabilistic Neural Network (PNN), the Bayesian and the Support Vector Machine (SVM) classifiers were tested. Also, the software provides a majority vote classifier combination option.

Feature selection: the exhaustive search (EXS) method was applied.

Evaluation: the leave-one-out, the cross validation and the self-consistency methods were utilized. For the estimation of the generalization performance of the system to unknown data, the external cross validation method was applied.

Parallel Processing: EXS is limited by serious computational and memory limitations. For bypassing these limitations, the CADMAMMO's pattern recognition system was designed on a Graphics Processing Unit (GPU) framework and it is capable of updating its structure in whenever a new verified case is uploaded on its repository.

Results-Discussion

- Results show that the most important single category of features is the textural features on the microscopy images with 87.3% prediction rate.
- When features from all families were concurrently tested, prediction accuracy was boosted up to 88.7% using a combination of five (5) textural, architectural and radiological features.

  These features were:

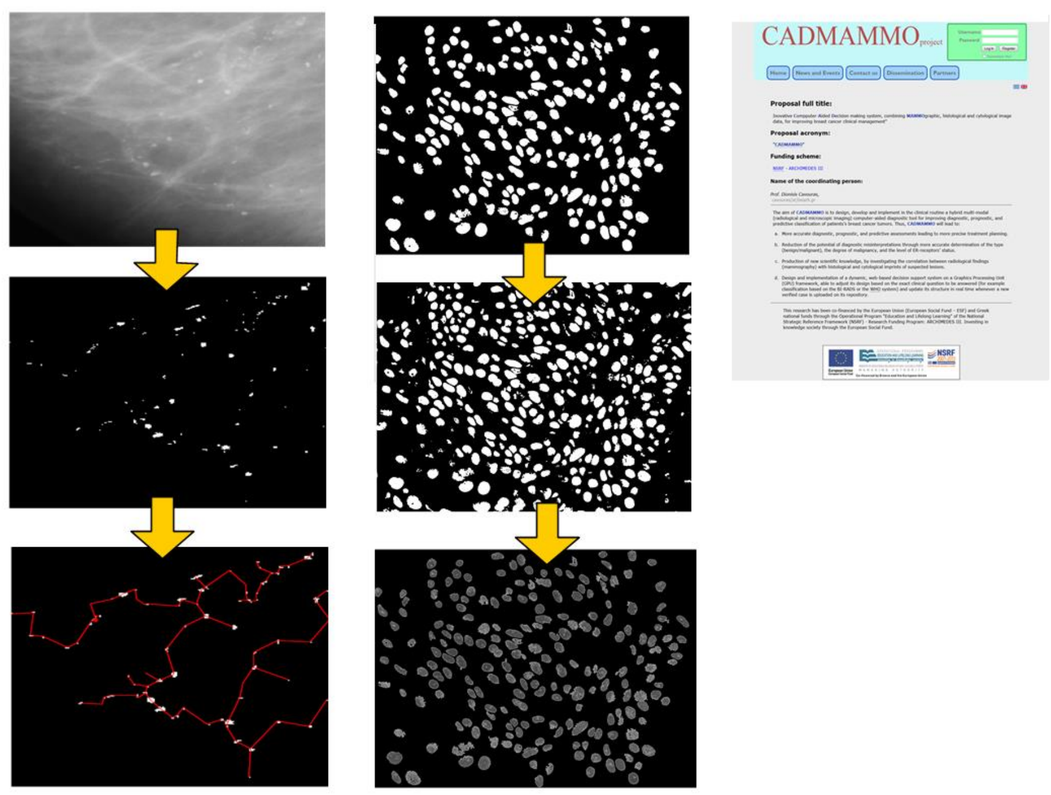
  1/ Mean Value of the grey level inside nuclei computed to estimate the average intensity over all segmented nuclei for each case,

  2/ Information Measure of Correlation-2 (average) that is a relative measure of texture 's entropy that tends to rise along image coarseness,

  3/ Maximal Correlation Coefficient (range) is a measure of homogeneity that tends to rise in inhomogeneous images,

  4/ Standard deviation of Minimal Spanning Tree is an nuclei topology measure that expresses the variance of nuclei vicinity and tends to decrease with nuclei stacking, plausibly caused by nuclei proliferation in high grade cases, and

  5/ mCa/lesion size; microcalcifications are an early stage sign of a potential dangerous lesion. In the presence of a lesion it implies information regarding the degree of abnormality.
- The combination of microcalcifications with the lesion's size proved significant in discriminating low from high grade breast cancer cases.



Conclusions

CADMAMMO services will provide a set of integrated, open, and interoperable tools of computer-assisted manipulation of radiological and microscopy images, contributing to more reliable estimations, measurements, and, ultimately, decisions regarding interpretation of imaging findings for medical professionals. In this way, diagnostic, prognostic, and predictive decisions will be improved to the benefit of patients, with great impact on social communities.

Acknowledgements

This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: ARCHIMEDES III. Investing in knowledge society through the European Social Fund.