A web-accessible mammographic image database dedicated to combined training and evaluation of radiologists and machines

Zinon C. Antoniou, Georgia P. Giannakopoulou, Ioannis I. Andreadis, Konstantina S. Nikita, Panos A. Ligomenides, and George M. Spyrou

***Abstract*—We designed and implemented a web-accessible database entitled MIRaCLe DB (Mammography Image reading for Radiologists' and Computers' Learning Database) that composes a dynamic repository for machines and radiologists training and evaluation. Up to now, 204 mammograms have been collected from 196 patients; they have been classified according to the type of the lesions, the type of the biopsy finding, the type of the mammography finding and the available CADs performance. A user can access the database in two modes: (a) for classification software evaluation and (b) for radiologists' education evaluation. In the mode for classification software evaluation, the user has the ability to query the database and retrieve cases with certain characteristics and certain difficulty. Also, there is the possibility to download the existing cases in order to facilitate the evaluation of a new classifier. In the other mode, the user (radiologist) can be trained in real time through a sequence of presentations and furthermore can be evaluated through different evaluating scenarios. At the duration of evaluation, the user can examine the mammography images through a web-based digital magnifier and process the corresponding image in real time. MIRaCLe DB is the first database that combines the machine and human needs for training and evaluation in mammographic image reading.**

***Index Terms*—Computer-aided diagnosis, radiologist evaluation, machine evaluation, mammography.**

I. Introduction

Breast cancer unfolds rapidly as a leading cause of death among women worldwide [1]. Early detection is the best way of defense against the breast cancer playing a momentous role in increasing the survival rate and diminish the emotional burden to the patients. Mammography is currently the best choice for detecting breast cancer at very early stage [2]. Traditionally, mammograms can be read by radiologists. However, Computer-Aided Diagnosis systems (CAD) for mammography have been already developed and used to enhance the validity in diagnosis providing a second opinion to the radiologists [3-12].

Manuscript received July 10, 2009.

Z. C. Antoniou, I. I. Andreadis and K. S. Nikita are with the National Technical University of Athens (e-mail: el04607@mail.ntua.gr, giannislavr@yahoo.com , knikita@cc.ece.ntua.gr).

G. P. Giannakopoulou, is with the University of Athens, School of Medicine.

P. A. Ligomenides and G. M. Spyrou are with the Academy of Athens, and the Academy of Athens Biomedical Research Foundation, Soranou Efessiou 4, 115 27 Athens, Greece (Corresponding Author: e-mail: gspyrou@bioacademy.gr).

The existence of mammography image databases is necessary in order to provide useful reference data sets for training and evaluation [13-20]. However, to our knowledge, there is not any platform dedicated to mammography that combines data suitable for machine and human reading evaluation. MIRaCLe DB (Mammography Image reading for Radiologists' and Computers' Learning Database) has been designed and implemented according to this trend. It composes a dynamic repository for machines and radiologists training and evaluation. The existing cases have been classified according to the type of the lesions, the type of the biopsy finding, the type of the mammography finding and the available CADs performance [21-23]. A user can access the database in two modes: (a) for classification software evaluation and (b) for radiologists' education evaluation. Such platform gives the possibility of performing objective benchmarking among (i) classification algorithms, (ii) radiologists performance and (iii) human-machine classification performance.

II. Material and Methods

A. Cases

Currently, MIRaCLe DB contains 204 mammography images collected from 196 patients from a Greek University Hospital. However it's expected to host a large number of cases from the collaboration consortium the Academy of Athens has started to build.

Each mammography image is linked to a set of respective information. This comprise the institution that it's taken from, the type of breast – left or right -, the biopsy result, the mammogram findings, the BIRADS number given by an expert, the dimensions of ROI –Region Of Interest- image, the ROI image and the CAD score (if any). Due to ethical constraints, all private information is excluded from the database.

All the candidate cases pass through a filtering process to attain high level of data quality, consistency and integrity.

B. Database Implementation and Use

The graphical user interface (GUI) of MIRaCLe DB is implemented using PHP scripting language. With the assistance of Cascading Style Sheet (CSS) language, we describe the presentation of some parts of web pages, like list tables and charts. On the background, there is a local database in MySQL to maintain our data, which it's

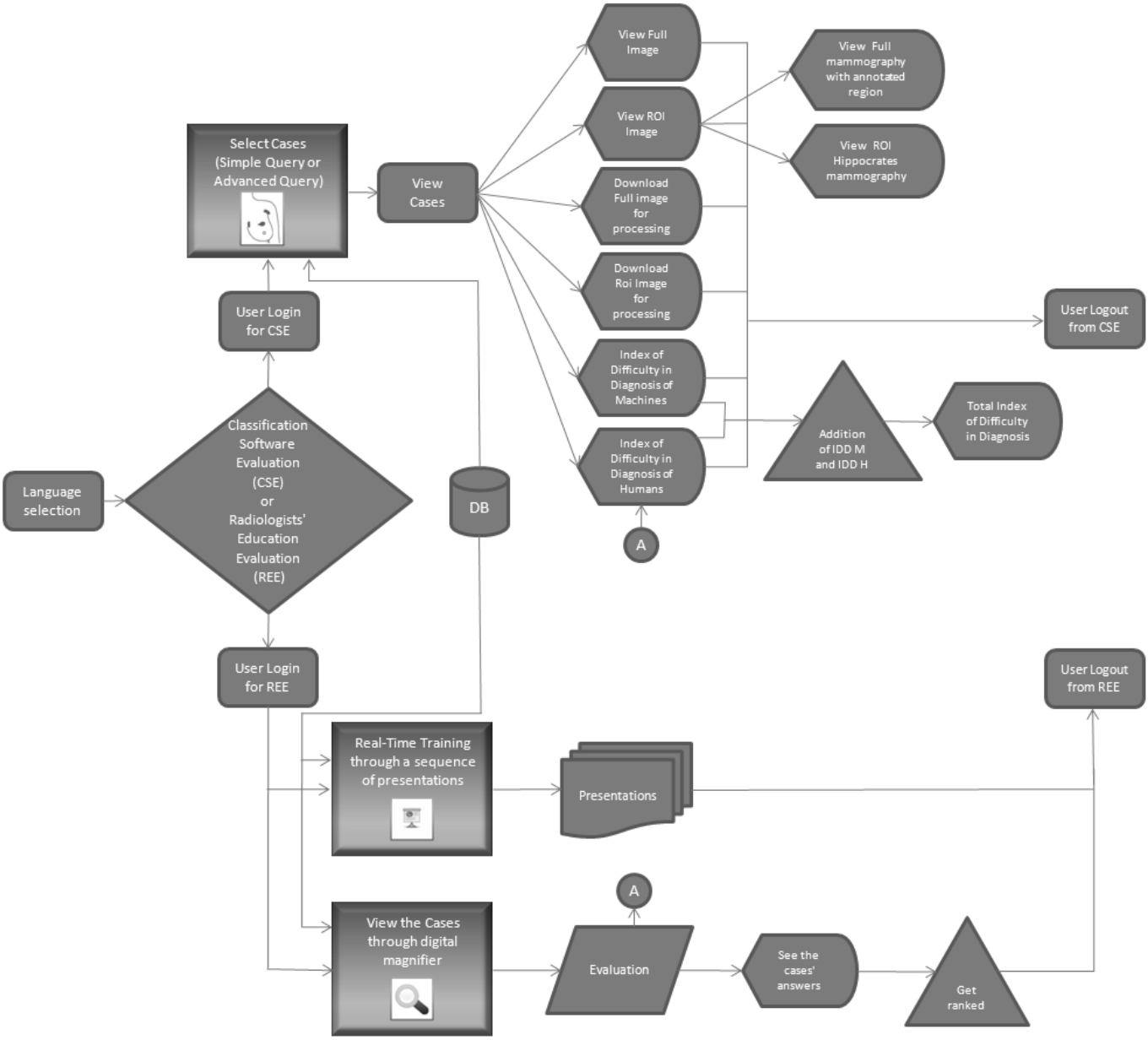

Fig 1. A flowchart with all the available tools of MIRaCLe DB.

carefully designed and implemented.

Besides, we used JavaScript scripting language to develop the digital magnifier tool and create an image viewer in a web page. Furthermore, we use a free flash tool - authorPOINT's Lite flash presentation, created by authorGen Technologies [24] - to present the training presentations for radiologists.

The only requirement for a user to have access to MIRaCLe DB is a personal computer with a web access.

Every new user must register for free in one of the two available services, either for Classification Software Evaluation or for Radiologists' Education Evaluation. Directly after the registration, the new user can browse through the database, using all the available tools described on figure 1. At this point, it's important to mention, that a user who registers only in one of the two available services, will have access only on that service. Just in case he/she wants to have access to the other service, he/she must register in that service respectively.

C. Classification Software Evaluation

Since the user is logged in Classification Software Evaluation, he/she is able to ramble through that part of the database. Observing the main menu, a user can see a list of the available functions of the database.

Firstly, he/she can see a list of Cases, presenting on charts. In addition, he/she can browse all the cases with statistics in a list mode. For each case listed on the table, a user is able to view the mammography image – full or ROI (Region Of Interest) image- and download it directly to his/her personal computer and process it with his/her CAD algorithm if he/she desires to verify its effectiveness and compare its results with other CAD algorithm results stored in database . However, if a user selects to view a ROI image, he/she will be able to view and download either the full mammography image with annotated region or the preprocessed ROI image having the corresponding finding (e.g. microcalcifications)

revealed.

Withal, each user has the capability to place a simple or an advanced query, using the proper search mode respectively.

When radiologists are in the midst of a reading session of a mammography image, the simple query, is offered for an individual keyword search, that would help them to retrieve quickly cases' information that are similar to their case.

Fig 2. Advanced query.

On the other hand, the advanced query, as seen on figure 2 above, is a considerable tool in our database, because a user is able to achieve complex queries and retrieve a specific result. For example, the user can retrieve all the images that were diagnosed as BIRADS 3 but they were malignant and the related finding was microcalcifications.

Besides, a user can make a more composite query and retrieve all images that although benign, they were diagnosed as BIRADS 4B and the CAD system gave a positive result. Then, the system will present all the images where both the radiologist and the CAD gave wrong diagnosis. Similar to this query, a user can retrieve all images that were benign and diagnosed as BIRADS 3 while the CAD system gave a positive result or all images that were benign and diagnosed as BIRADS 4B while the CAD system gave a negative result. The result of these queries will present which images are difficult to discern as benign or malignant by radiologists or by a CAD system respectively.

Still, users can browse over the evaluated CADs that are stored in database and compare their performance features, such as sensitivity, specificity and other indices of accuracy.

Likewise, while a user is logged in Classification Software Evaluation, after he/she performs a full download of a group of images and subsequently evaluation of a new CAD algorithm, he/she can give feedback to MIRaCLe DB through a certain template. This function intends to help us to materialize two of our goals. The one is to exhort people who implement CAD systems' algorithms and techniques to improve their effectiveness, due to spirit of competition. The other one is to enrich the database content and keep it updated.

D. Radiologists' Education Evaluation

While the user is logged in Radiologists' Education Evaluation, he/she is able to ramble through that part of the database. Observing the main menu, a user can see a list of the available functions.

Under this part of MIRaCLe DB there are two sections. In the first one, young radiologists have the ability to train themselves in real time, through a sequence of presentations created by experts.

In the other one, radiologists have the ability to be evaluated. For each evaluating case, they can see the mammography image through a digital magnifier tool, as seen on figure 3 below. For each case, radiologists would be asked to choose the correct diagnosis from a list, that corresponds to the mammography image findings. At the end of the evaluation, they would get their score and esteem their diagnosis abilities.

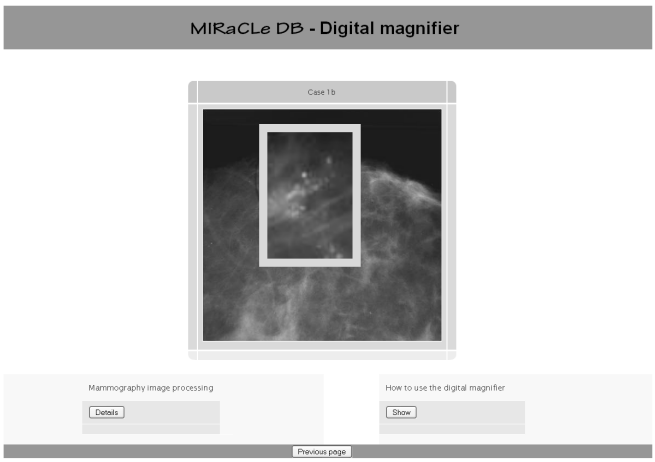

Fig 3. Digital magnifier in radiologists' evaluation.

Each user can be evaluated only once with the same account. However, if they wish, they can register once again with the same personal information, using different username and password this time.

Training and evaluation of the radiologists will raise their awareness and help them to find out any weakness.

III. DISCUSSION

We have developed a secured database for storage and retrieval of mammography images with a set of respective information for evaluation of classifiers, machines and human. MIRaCLe DB is accessible through any computer with web access.

After the user ID and password is verified either in Classification Software Evaluation or in Radiologists' Education Evaluation, the system will allow a user to view all the available cases stored in database or train or/and evaluate himself/herself respectively. This aspect may constitutes a challenge for the valid evaluation of the effectiveness of the software classifiers as well as that of the radiologists.

Besides, MIRaCLe DB aspires to satisfy the radiologists'

needs that work in provincial hospitals to deal with plenty of mammography images and biopsy results, and compare them with the case that they are examining.

Further work is in the pipeline as for the functions of MIRaCLe DB. According to the data stored in database, both from machines and humans evaluation, a user will have the ability to browse over difficulty driven cases based on machine and human performance on each case. The index of difficulty is going to constitute a motivation for the researchers, to implement more efficient algorithms and techniques for CAD systems and for radiologists to improve their discrimination over mammography images.

REFERENCES

[1] World Health Organization. (2009, February) *Fact sheet.* No.297, [Online]. Available: http://www.who.int/mediacentre/factsheets/fs297/en/index.html

[2] M. Reddy, and R. Given-Wilson, "Screening of breast cancer", *Women's Health Medicine*, vol. 3, issue 1, pp. 22-27, January-February 2006.

[3] J. Y. Lo, M. Gavrielides, M. K. Markey, and J. L. Jesneck, "Computer-aided classification of breast microcalcification clusters: Merging of features from image processing and radiologists", *Proceedings of SPIE*, vol. 5032, pp. 882-889, 2003.

[4] T. W. Freer, and M. J. Ulissey, "Screening Mammography with Computer-aided Detection: Prospective Study of 12,860 Patients in a Community Breast Center", *Radiology*, vol. 220, pp. 781-786, 2001.

[5] B. Zheng, L. A. Hardesty, W. R. Poller, J. H. Sumkin, and S. Golla, "Mammography with Computer-aided Detection: Reproducibility Assessment - Initial Experience", *Radiology*, 2003, vol. 228, pp. 58-62.

[6] Y. Jiang, R. M. Nishikawa, R. A. Schmidt, C. E. Metz, M. L. Giger, and K. Doi, " Improving Breast Cancer Diagnosis with Computer-aided Diagnosis", *Acad Radiol*, vol. 6, pp. 22-33, 1999.

[7] P. Deloglu, M. E. Fantacci, A. P. Martinez, A. Retico, A. Stefanini, and A. Tata, "A scalable system for microcalcification cluster automated detection in a distributed mammographic database", in *Nuclear Science Symposium Conference Record*, vol. 3, October 2005.

[8] A. Lauria, "GPCALMA: Implementation in Italian hospitals of a computer aided detection system for breast lesions by mammography examination", *Physica Medica*, vol. 25, pp. 58-72, 2009.

[9] I. Abdel-Qader, and F. Abu-Amara, "A Computer-Aided Diagnosis System for Breast Cancer Using Independent Component Analysis and Fuzzy Classifier", *Modeling and Simulation in Engineering*, vol. 2008, 2008.

[10] M. K. Markey, J. Y. Lo, G. D. Tourassi, and C.E. Floyd Jr., "Self-organizing map for cluster analysis of a breast cancer database", *Artificial Intelligence in Medicine*, vol. 27, pp. 113-127, 2003.

[11] A. Lauria, M. E. Fantacci, U. Bottigli, P. Delogu, F. Fauci, B. Golosio, P.L. Indovina, G. L. Masala, P. Oliva, R. Palmiero, G. Raso, S. Stumbo, and S. Tangaro, "Diagnostic performance of radiologists with and without different CAD systems for mammography", *Proceedings of SPIE*, vol. 5034, pp. 244-248, 2003.

[12] S. M. Astley and F. J. Gilbert, "Computer-aided detection in mammography", *Clinical Radiology*, vol. 59, pp. 390-399, 2004.

[13] G. J. Becker, "The Potential Power of Reference Image Databases and Oncologic Imaging", *Journal of the American College of Radiology*, vol.2, issue 7, pp.565-567, July 2005.

[14] D. Tahmoush, and H. Samet, "A Web Database for Computer-Aided Detection and Diagnosis of Medical Images", *Lecture Notes in Computer Science*, vol. 5116, pp. 265-272, 2008.

[15] R. Warren, A. E. Solomonides, C. del Frate, I. Warsi, J. Ding, M. Odeh, R. McClatchey, C. Tromans, M. Brady, R. Highnam, M. Cordell, F. Estrella, M. Bazzochi, and S. R. Amendolia, "Mammogrid – a prototype distributed mammographic database for Europe", *Clinical Radiology*, vol. 72, pp. 1044-1051, 2007.

[16] A. Lauria, R. Massafra, S. Sonia Tangaro, R. Bellotti, M. Evelina Fantacci, P. Deloglu, E. L. Torres, P. Cerello, F. Fauci, R. Magro, and U. Bottigli, "GPCALMA: An Italian Mammographic Database of Digitized Images for Research", *Lecture Notes in Computer Science*, vol. 4046, pp. 384-391, 2006.

[17] M. L. Chang, C. P. Chuang, T. L. Kuo, C. C. Wu, G. C. Hsu, C. P. Huang, Y. L. Chang, and S. K. Lee, "A Pilot/Learning/Training/Teaching System of Mammography Screening for Radiologists", in *Asia Pacific Association of Medical Informatics Conference*, October, 2006.

[18] L. P. Clarke, and B. Croft, "Development of Public Resources to Support Quantitative Imaging Methods in Cancer", *Acad Radiol*, vol. 14, pp. 1438-1440, 2007.

[19] M. Wu, Y. Zheng, M. North, and E. Pisano, "NLM Tele-Educational Application for Radiologists to Interpret Mammography", in 2002 *Proceedings of the AMIA Symposium*, pp. 909-913.

[20] M. Heath, K. Bowyer, D. Kopans, R. Moore and, P. Kegelmeyer Jr, "The Digital Database for Screening Mammography", *Proceedings of the Fifth International Workshop on Digital Mammography*, pp. 212-218, 2001.

[21] A. Frigas, S. Kapsimalakou, G. Spyrou, K. Koufopoulos, S. Vassilaros, A. Chatzimichael, J. Mantas, and P. Ligomenides, "Evaluation of a Breast Cancer Computer Aided Diagnosis System", in *Stud Health Technol Inform*, vol. 124, pp. 631-636, 2006.

[22] G. Spyrou, S. Kapsimalakou A. Frigas K. Koufopoulos S. Vassilaros, and P. Ligomenides ""Hippocrates-mst": A prototype for Computer-Aided Microcalcification Analysis and Risk Assessment for Breast Cancer", *Medical & Biological Engineering & Computing*, vol. 44, issue 11, pp. 1007-1015, November 2006.

[23] A. Frigas, G. Spyrou, P. Ligomenides, M. Diomidous, and J. Mantas, "Investigating the Classification Robustness of a Computer Aided Diagnosis System Using Mammographic Images", *Acta Informatica Medica*, vol. 17, issue 2, pp. 64-66, 2009.

[24] authorPOINT Lite, [Online]. Available: http://www.authorgen.com/authorpoint-lite-free/powerpoint-to-flash-converter.aspx