Research Methodology: An example in a Real Project

Noel Pérez

Laboratory of Optics and Experimental Mechanics, Instituto de Engenharia Mecânica e Gestão Industrial.
nperez@inegi.up.pt

Abstract. The research methodology defines what the activity of research is, how to proceed, how to measure progress, and what constitutes success. It provides us an advancement of wealth of human knowledge, tools of the trade to carry out research, tools to look at things in life objectively; develops a critical and scientific attitude, disciplined thinking to observe objectively (scientific deduction and inductive thinking); skills of research particularly in the 'age of information'. Also it defines the way in which the data are collected in a research project. In this paper it presents two components of the research methodology from a real project; the theorical design and framework respectively.

Keywords: Research methodology, example of research methodology, theorical framework, theorical design.

1 Introduction

The research methodology defines what the activity of research is, how to proceed, how to measure progress, and what constitutes success. It provides us an advancement of wealth of human knowledge, tools of the trade to carry out research, tools to look at things in life objectively; develops a critical and scientific attitude, disciplined thinking to observe objectively (scientific deduction and inductive thinking); skills of research particularly in the 'age of information'.

The research methodology is a science that studying how research is done scientifically. It is the way to systematically solve the research problem by logically adopting various steps. Also it defines the way in which the data are collected in a research project.

1.1 Study case

According to the World Health Organization (WHO) breast cancer is the most common cancer suffered by women in the world, which during the last two decades has increased the women mortality in developing countries. Mammography is the best method used for screening; it is a test producing no inconvenience and with small diagnostic doubts of breast cancer since the preclinical phase [1]. The role of screening mammography in the battle against breast cancer is well established;

women with malignancies detected at an early stage have a significantly better prognosis. However, it is also recognized that the diagnostic interpretation of mammograms continues to be challenging for radiologists with a documented 20% false negative rate [2]. The clinical significance of early breast cancer diagnosis and the higher than desired false-negative rate of screening mammography have motivated the development of computer-aided detection/classification (CAD) systems for decision support. These systems typically involve a hierarchical approach, first applying elaborated image preprocessing steps to enhance suspicious structures in the image and then employing morphologic and textural analysis to better classify these structures between true abnormalities and false positives [2-4]. We made a detailed review of techniques for mammographic image analysis and related CAD systems. This review included methods and techniques from different mammography images sources such as conventional screen film mammography and full-field digital mammography [1-3, 5-9] to ultrasound (US), magnetic resonance imaging (MRI), and computed tomography (CT) images [10-13]. Although true clinical impact of CAD systems is often debated, the scientific community continues to work toward improving the diagnostic performance and clinical integration of CAD technology. For this reason, we consider that reliable CAD systems for automated detection/classification of pathological lesions (PL) will be very useful and helpful to supply a valuable "second opinion" to medical personnel.

This project is focused to develop novel methods and algorithms to improve the following fields: image contrast enhancing, accurate PL segmentation, features vectors extraction and the classifiers accuracy to reduce classification errors.

Our intention is to build a more robust computerized framework and implemented it on an appropriated distributed computing (GRID) environment to expand their possibilities to medical communities, for creating, hosting and managing GRID-based mammography digital repositories. This framework will facilitate the massive study and analysis of breast cancer in mammography images and we consider it the needed support to design, develop and evaluate more reliable and robust CAD systems.

2 Theorical Framework

State of the art of "Development and Evaluation of Mammography Images Analysis Algorithms in GRID Environment" base on digital image processing, pattern recognition and artificial intelligence techniques. Some examples of developed methods with interesting results, in which is inspired this project proposal are outlined below:

- An approach to compute morphology/texture features of breast lesions, which are associated with lesions phenotype appearance on MRI, were used for diagnostic prediction. Six features, including compactness, normalized radial length entropy, volume, gray level entropy, gray level sum average, and homogeneity were selected by an Artificial Neural Network (ANN) using leave-one-out cross validation method. The area under the receiver-operating characteristic (ROC [4]) curve was 0.86. When dividing the database into half training and half validation set, a classifier of five features selected in the half training set achieved an area under the curve of 0.82 in the other half validation set, demonstrating that these features could be used by an ANN to form a classifier for differential diagnosis [13].

- A method to extract automatically identified image possible PL and produce a set of selected features (mathematic descriptors), which are merged into an estimate of the probability of malignancy using a Bayesian ANN classifier. This method was validate on seven hundred thirty-nine full-field digital mammography (FFDM) images, which contained 287 biopsy-proven breast mass PL, of which 148 lesions were malignant and 139 lesions were benign. Lesion margins were delineated by an expert breast radiologist and were used as the truth for lesion-segmentation evaluation. Performance of the analyses was evaluated at various stages of the conversion using ROC analysis. An area under the ROC curve value of 0.81 was obtained in the task of distinguishing between malignant and benign mass lesions in a round-robin by case evaluation on the entire FFDM dataset [8].

- A CAD system that allows to select manually possible PL and produce automatically a features vector (composed by: PL area, average of PL intensities levels (brightness), PL shape and PL elongation), which is used by a trained ANN to diagnose six classes of mammography PL: calcifications, well-defined/circumscribed masses, spiculated masses, ill-defined masses, architectural distortions and asymmetries) as benign or malignant tissues. This system was validated on the Mammographic Image Analysis Society (MIAS) database, with a representative dataset formed by 100 images selected randomly (including examples of all PLs classes). The system performance was evaluated with different ANN models and confirmed successfully in the: feedforward backpropagation (FB) and generalized regression (GR) obtaining a classification result of 94.0% and 80.0% of true positives respectively [3].

Despite the image input source, we consider that a suitable combination of digital image processing, pattern recognition and artificial intelligence techniques is the key to expand the mammography CAD performance.

3 Theorical design categories

3.1 Scientific problem

Insufficiency in mammography images analysis techniques on GRID environment platform used in CETA-CIEMAT.

3.2 Research object

Mammography images analysis process

3.3 Research objective

Development a set of mammography images analysis algorithms for a GRID environment

3.4 Research field

Digital image processing, pattern recognition and artificial intelligence techniques

3.5 Scientific hypothesis

If it develops a set of mammography images analysis algorithms based on digital image processing, pattern recognition and artificial intelligence techniques, we can reduce the insufficiency in mammography images analysis techniques on GRID environment platform used in CETA-CIEMAT.

3.5.1 Independent variable

Set of mammography images analysis algorithms based on digital image processing, pattern recognition and artificial intelligence techniques

3.5.2 Dependent variable

Reduce the insufficiency in mammography images analysis techniques on GRID environment platform used in CETA-CIEMAT.

3.6 Research task

3.6.1 Facto-perceptible stage

- Determination of the historical development of the digital image processing, pattern recognition and artificial intelligence techniques.
- Gnoseology Characterization of the mammography images analysis process.
- Gnoseology Characterization of the digital image processing, pattern recognition and artificial intelligence techniques
- Characterization of the current state of mammography images analysis process.

3.6.2 Theorical preparation stage

- Design of mammography images analysis algorithms based on digital image processing, pattern recognition and artificial intelligence techniques.
- Algorithms implementation.

3.6.3 Application stage

- Validation of the results obtained by developed algorithms in mammography images analysis process.

4 Methodological Design

4.1 Research type and general goal

- The proposed research is applied, descriptive, experimental and longitudinal.
- The proposed research is developed from quantitative point of view.

4.2 Population and sample

- The population of the proposed research is formed by 322 mammography images from the miniMIAS public database and the specialist personnel team.
- The population represents a sample of possible anomalies (pathological lesions) that could be appearing in the algorithms evaluation process.

4.3 Methods and Techniques

In this research we will use the following methods to accomplish the proposed tasks:

- The historic-logic method was used to determine the previous research about of mammography images analysis process.
- The synthetic-analytic method was used to process the obtained information from the specialist personnel (expert's opinion).

4.4 Statistical processing of information

- Effectiveness percent analysis of the developed algorithms.
- The nonparametric method (Mann-Whitney U Test [14]) is applied to determine significant differences among the used set of data.

4.5 Work schedule

Stage	Task
Facto-perceptible	- Adjustment of the research redesign. - Selection and location of the bibliography. - Accuracy of historical antecedents. o Literature search about research field and object. o Wording and style revision of the historical antecedents. - Presentation of the historical antecedents to the scientific committee. - Construction of contextual antecedents. - Construction of reference antecedents.
Elaboration	- Study of the programming language where will carry out the algorithms implementation. - Algorithms implementation.
Application	- Assessment and corroboration of the results. - Construction of the conclusions.

	- Presentation of the developed algorithms to the scientific committee. - Re-evaluation of the results. - Completion of the research.

5 Conclusion

The research methodology helps to learn how to use libraries and other information resources, enables critical evaluation of literature; develops special interests and skills. Helps to understand attitude of others and creates awareness of special needs of research process. Describes and analyze methods, throw light on their limitations and resources, clarify their presupposition and consequences, relating their potentialities to the twilight zone at the "frontiers of knowledge".

References

1 Y. López, A. Novoa, M. Guevara, and A. Silva, “Breast Cancer Diagnosis Based on a Suitable Combination of Deformable Models and Artificial Neural Networks Techniques,” in Progress in Pattern Recognition, Image Analysis and Applications, 2008, pp. 803-811.

2 G.D. Tourassi, R. Ike Iii, S. Singh, and B. Harrawood, “Evaluating the Effect of Image Preprocessing on an Information-Theoretic CAD System in Mammography,” Academic Radiology, vol. 15, (no. 5), pp. 626-634, 2008.

3 Y. López, A. Novoa, M. Guevara, N. Quintana, and A. Silva, “Computer Aided Diagnosis System to Detect Breast Cancer Pathological Lesions,” in Progress in Pattern Recognition, Image Analysis and Applications, 2008, pp. 453-460.

4 R.F. Wagner, C.E. Metz, and G. Campbell, “Assessment of Medical Imaging Systems and Computer Aids: A Tutorial Review,” Academic Radiology, vol. 14, (no. 6), pp. 723-748, 2007.

5 W. Qian, D. Song, M. Lei, R. Sankar, and E. Eikman, “Computer-Aided Mass Detection Based on Ipsilateral Multiview Mammograms,” Academic Radiology, vol. 14, (no. 5), pp. 530-538, 2007.

6 Z. Jeffrey Zhi-jie, L. Liang, and X. Yinfu, “Towards automated mammograph image analysis,” in Book Towards automated mammograph image analysis, Series Towards automated mammograph image analysis, Editor ed.^eds., City, 2005, pp. 6 pp.

7 K. Byrd, Z. Jianchao, and M. Chouikha, “Performance assessment of mammography image segmentation algorithms,” in Book Performance assessment of mammography image segmentation algorithms, Series Performance assessment of mammography image segmentation algorithms, Editor ed.^eds., City, 2005, pp. 6 pp.-157.

8 H. Li, M.L. Giger, Y. Yuan, W. Chen, K. Horsch, L. Lan, A.R. Jamieson, C.A. Sennett, and S.A. Jansen, “Evaluation of Computer-aided Diagnosis on a Large Clinical Full-field Digital Mammographic Dataset,” Academic Radiology, vol. 15, (no. 11), pp. 1437-1445, 2008.

9 L. Jiang, E. Song, X. Xu, G. Ma, and B. Zheng, “Automated Detection of Breast Mass Spiculation Levels and Evaluation of Scheme Performance,” Academic Radiology, vol. 15, (no. 12), pp. 1534-1544, 2008.

10 W.-J. Wu and W.K. Moon, “Ultrasound Breast Tumor Image Computer-Aided Diagnosis With Texture and Morphological Features,” Academic Radiology, vol. 15, (no. 7), pp. 873-880, 2008.

11 W. DeMartini, C. Lehman, and S. Partridge, “Breast MRI for Cancer Detection and Characterization: A Review of Evidence-Based Clinical Applications,” Academic Radiology, vol. 15, (no. 4), pp. 408-416, 2008.

12 X. Liang, Q. Zhang, C. Li, S.R. Grobmyer, L.L. Fajardo, and H. Jiang, “Phase-Contrast Diffuse Optical Tomography: Pilot Results in the Breast,” Academic Radiology, vol. 15, (no. 7), pp. 859-866, 2008.

13 K. Nie, J.-H. Chen, H.J. Yu, Y. Chu, O. Nalcioglu, and M.-Y. Su, “Quantitative Analysis of Lesion Morphology and Texture Features for Diagnostic Prediction in Breast MRI,” Academic Radiology, vol. 15, (no. 12), pp. 1513-1525, 2008.

14 William C. Black , Robert F. Nease, Jr , and Anna N. A. Tosteson, “Perceptions of Breast Cancer Risk and Screening Effectiveness in Women Younger Than 50 Years of Age”. J. Natl. Cancer Inst. 87: 720-731.