APPLICATION OF DATA MINING TECHNIQUES FOR MAMMOGRAM CLASSIFICATION

Milos Radovic, Marina Djokovic, Aleksandar Peulic, Nenad Filipovic
University of Kragujevac, Serbia

Introduction

One of the leading causes of cancer death among women is breast cancer. In our work we aim at proposing a prototype of a medical expert system (based on data mining techniques) that could significantly aid medical experts to detect breast cancer. Normal breast images and breast image with masses (total 200 images) used as input images in this study are taken from the Clinical Center in Kragujevac.

Methods

This paper presents the CAD (computer aided diagnosis) system for the detection of normal and abnormal pattern in the breast. The proposed system consists of four major steps: the image preprocessing, the feature extraction, the feature selection and the classification process that classifies mammogram into normal (without tumor) and abnormal (with tumor) pattern.

After removing noise from mammogram using the Discrete Wavelet Transformation (DWT), first is selected the region of interest (ROI). By identifying the boundary of the breast, it is possible to remove any artifact present outside the breast area, such as patient markings (Figure 2).

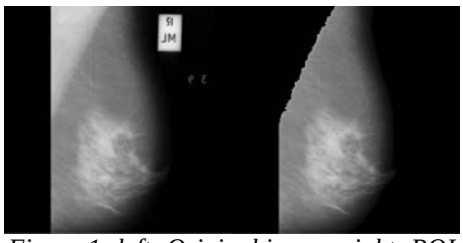

Figure 1: left: Original image, right: ROI

Then, a total of 20 GLCM features are extracted from the ROI, which were used as inputs for different data mining algorithms capable to deal with classification problems. Texture features is considered to be one of the widely used tool in masses detection in digital mammogram [Pradeep, 2012].

In order to improve classification results, different feature selection algorithms have been used to select the most relevant features. For classification of mammograms we used several data mining algorithms: Support Vector Machine classifier, Naive Bayes classifier, K-Nearest Neighbor classifier, logistic regression, decision trees, random forest and neural network.

The basic idea of procedure for automatic medical image classification, when applied to images, consists of three steps: (1) Texture Feature Extraction, (2) Feature Selection and (3) Classification. This procedure is shown in Figure 2.

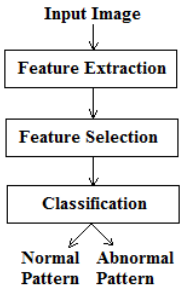

Figure 2: Proposed methodology for mammographic image classification

Results and discussion

This paper shows that advanced techniques of image processing and the masses detection are useful in computer aided diagnosis. Many missed radiologist diagnoses can be attributed to human factors such as subjective or varying decision criteria, distraction by other image features or large number of images to be inspected. Therefore, computer-aided diagnosis is an important research area. The methods like one presented in this paper could assist the radiologist and improve the accuracy of detection.

Detection is done based on textural descriptors obtained from features extraction process. Results show that few of proposed data mining algorithms are able to deal with the problem of mammogram classification. This approach has potential for further development because of its simplicity that will motivate real-time breast cancer diagnosis in providing a second opinion to radiologists.

References

Pradeep et al, Computer Engineering and Intelligent Systems, 3(4):37- 46, 2012.