International Conference on Computational Modeling and Security (CMS 2016)

Image Mining using Association Rule for Medical Image dataset

Jyoti Deshmukh[a,]* and Udhav Bhosle[b]

[a,b]*Department of Electronics and Telecommunication Engineering, Rajiv Gandhi Institute of Technology, Mumbai-400 053, India*

Abstract

The concept of data mining for discovering frequent image patterns in mammogram images using association rule is presented. Proposed method works in two phases. First phase is segmentation of digital mammogram to find region of interest (ROI). It consists of median filtering for noise removal, morphological processing for removing the background and suppressing artifacts, image enhancement techniques to improve image quality followed by region growing algorithm for complete removal of pectoral muscle. Second phase is image mining to find frequent image patterns present in mammogram images using Association rule. It consists of feature extraction, optimization by selecting most discriminating features among them, discretization of selected features and generation of transaction representation of input images. This is given as input to Apriori algorithm to generate association rules. The proposed method uses a new ESAR (Extraction of strong association rule) algorithm to obtain strong, effective and highly correlated association rules from the rules obtained using Apriori algorithm in previous step. Result shows that image mining is feasible and gives strong association rules. These association rules can be further used for effective diagnosis of mammogram images.

*Keywords :*Image mining*, Association rules, Support, Confidence, Region of interest, Correlation measures;

1. Introduction

Every day, huge amount of satellite images, medical images and digital photographs are generated. Useful information can be revealed to the human users by analysing these images. As a result, there is a demand of image

* Corresponding author. Tel.: +91-9892724542; Fax:022-26707026
  E-mail address: jyoti.deshmukh@mctrgit.ac.in, udhav.bhosle@mctrgit.ac.in

mining systems, which can automatically review semantically meaningful information (knowledge) from huge amount of image data. A very large number of mammograms are generated every day in hospitals and medical centres. Process of analyzing and diagnosing mammogram has become critical. As a result, there is a need of computer aided diagnosis (CAD) system to help the physician's task. In this proposed method, we find association rules, which represent frequent pattern that occurs together in similar type of images i.e. benign or malignant mammogram images and these rules, can be used further for effective diagnosis of mammogram. In proposed method we use most discriminating features during mining process to get strong association rules. Our method makes the image mining algorithm faster, as it uses feature optimization by selecting most discriminating features and feature discretization. Proposed method works in two phases. First is segmentation of digital mammogram for finding Region of Interest (ROI) and second is finding frequent patterns in mammogram images using Apriori algorithm of Association Rule mining. We proposed a new ESAR algorithm for optimizing these generated association rules to get strong, effective and highly correlated association rules.

Agrawal et al. [1] discussed the problem of Association rule mining very first time. Beyer et al. [2] presented increasing the number of features to represent the image can create a problem. Thus to improve the discrimination accuracy, we have to keep the count of features as low as possible. Color distribution of image is expressed by histogram, but for breast lesions it provides poor identification. To differentiate mammogram lesions as malignant and benign, shape features can be used but it increases the computational complexity of the process. Textural variations in mammogram represent the differences in density of breast tissue. Properties of roughness, smoothness and regularity are described by texture feature.

One can query a database by using texture as a visual feature to retrieve similar patterns. The spatial arrangement of pixel intensities characterizes texture information [3][4]. Mammogram describes the tissue density differences and these differences are very important for analysis of mammogram. The differences in the density of breast tissue can be captured in a mammogram in the form of textural variations. [5] C. Ordonez et al. [6] introduced data mining for knowledge discovery in image database. They concentrate on the problem of finding association rules in 2-dimensional color images. Carson et.al. [7] presented image representation which provides a transformation from raw pixel data to a small set of localized coherent regions in color and texture space. Ji Zhang et al. [8] proposed various image mining research issues, frameworks used for image mining, current developments in image mining, state of the art techniques and systems. A framework for texture information of an image and achievement of higher retrieval efficiency than the shape features of an image is presented by Monika Sahu et al. [9]. Marcela Y. Ribeiro et al. [13] discussed a mammogram classification method based on association rule mining to improve the diagnosis of mammograms. This method produces both non-sensitive and sensitive association rules. For the diagnosis process, these non-sensitive association rules are not helpful. Also they are finding region of interest (ROI) of mammogram manually and then to these ROI, feature extraction techniques were applied.

Maria-Luiza Antonie et al. [14, 15, 16] proposed a mammogram classification method using association rules. Authors used Neural Network as a classifier and association rule mining as the data mining algorithm. Combination of a rough set theory along with association rules is used for mammogram clarification by Jiang Yun et al. [17]. Sumeet Dua et al. [18] presented weighted association rule based classification. It uses inter-class and intra-class weight of each association rule for classification. Jawad Nagi et al. [19] proposed a method using morphological processing and seeded region growing algorithm for automated breast profile segmentation.

Although many of the researchers have developed different techniques for mining of mammogram images to find strong and efficient association rules, still it is a challenging task. Hence we proposed a texture based image mining method using association rule for mammogram images. The rest of this paper is organized as follows; section 2 represents proposed image mining method. Section 3 represents Experimental result. Section 4 summarizes the conclusion and the future work.

2. Proposed image mining method

Figure 1 shows the block diagram of proposed method. It works in two phases. First is a segmentation of digital mammograms for finding region of interest (ROI) and second is finding frequent patterns in mammogram images using association rule mining. Each input mammogram image is associated with a keyword i.e. benign or malignant. Figure 2 shows algorithm for proposed method.