Biomedical Engineering: Applications, Basis and Communications, Vol. 24, No. 6 (2012) 513–524
DOI: 10.1142/S1016237212500470

ASSOCIATIVE CLASSIFICATION OF MAMMOGRAMS BASED ON PARALLEL MINING OF IMAGE BLOCKS

Mohsen Alavash Shooshtari*, Keivan Maghooli*,‡ and Kambiz Badie†

*Department of Biomedical Engineering, Science and Research Branch, Islamic Azad University
Hesarak, Tehran, Iran

†Research Institute for Information and Communication Technology, Tehran, Iran

‡k_maghooli@srbiau.ac.ir

Accepted 15 May 2012
Published 19 October 2012

ABSTRACT

One of the main objectives of data mining as a promising multidisciplinary field in computer science is to provide a classification model to be used for decision support purposes. In the medical imaging domain, mammograms classification is a difficult diagnostic task which calls for development of automated classification systems. Associative classification, as a special case of association rules mining, has been adopted in classification problems for years. In this paper, an associative classification framework based on parallel mining of image blocks is proposed to be used for mammograms discrimination. Indeed, association rules mining is applied to a commonly used mammography image database to classify digital mammograms into three categories, namely normal, benign and malign. In order to do so, first images are preprocessed and then features are extracted from non-overlapping image blocks and discretized for rule discovery. Association rules are then discovered through parallel mining of transactional databases which correspond to the image blocks, and finally are used within a unique decision-making scheme to predict the class of unknown samples. Finally, experiments are conducted to assess the effectiveness of the proposed framework. Results show that the proposed framework proved successful in terms of accuracy, precision, and recall, and suggest that the framework could be used as the core of any future associative classifier to support mammograms discrimination.

Keywords: Association rules; Image mining; Mammograms classification; Medical decision making.

INTRODUCTION

Data mining, as a promising multidisciplinary field in computer science, in its simplest notion refers to extracting knowledge from large amounts of data.[1] The common practice, hence, is to discover interesting patterns from different information repositories, including transactional, spatial, time-series, or multimedia databases. In practice, the two primary objectives of data mining tend to be prediction and description.[2] With respect to predictive mining, the aim is to build a model, expressed as an executable algorithm, which can be used to perform classification, prediction, estimation, or other similar tasks. Since the advent of data mining technology, many techniques have been proposed to

‡Corresponding author: Keivan Maghooli, Department of Biomedical Engineering, Science and Research Branch, Islamic Azad University, Poonak Square, Ashrafi Esfahani Boulevard, Hesarak, Tehran, Iran. Tel: +98(0)2144474321; Fax: +98(0)21444743 19; E-mail: k_maghooli@srbiau.ac.ir

implement such algorithms. One of these potential techniques is mining association rules, first introduced by Agrawal *et al.*[3] Roughly speaking, an association rule can be regarded as a relationship of the form $X \Rightarrow Y$, where X and Y are two separate sets of data items.[4] Given a set of transactions T (the transactional database), the association rules mining problem is to discover all association rules that have a *support* and a *confidence* greater than the user-specified minimum support and minimum confidence.[5] In this respect, support is the percentage of transactions containing both X and Y in the whole transactional database (so can be treated as the *rule coverage*), whereas confidence is the ratio of the number of transactions that contain X and Y over the number of transactions that contain X (so sometimes named as *accuracy*). The basic *a priori* algorithm introduced by Agrawal *et al.*[3] has been regarded as the most conventional association rules mining algorithm. Association rules are really no different from classification rules except that the former can predict any attribute, not just the class, and this gives them the freedom to predict combinations of attributes too.[6] Classification problems have adopted association rules for years (associative classification). Associative classification rules are learned and extracted from the available database in a supervised fashion. The most suitable rules are then selected to build an associative classification model.[5] Associative classification offers the following advantages over other classification rules generation techniques:

- Training of the classifier is generally much faster than other classification rules generation techniques such as decision tree.
- Training sets with high dimensionality can be handled very effectively.
- The resulting classifier is expressed as a set of rules which are easily understandable and simple to apply to unseen data.[7]

Association rules have been applied, for example, to texture and image classification as well as segmentation.[8−11] The statistical and structural information of an image are embedded in the local intensity variation patterns of the image regions. Association rules could be used to identify these frequently occurring patterns in the image and discover relationships that have significant discriminative power.

Applications of associative classification have been also extended to medical images. These images contain several correlated features, often addressing the detection of abnormalities, which when mined and exploited can lead to invaluable discriminative diagnosis.[12] For example, association rules mining has been previously used to classify CT scan brain images as well as ophthalmologic images, and the use of associative classifiers as an efficient technique to assist the diagnostic task has been suggested.[13−15] The result of such efforts could be used to support the development of computer-aided diagnosis (CAD) systems. These are automated systems that incorporate techniques such as image mining to assist the medical staff.[16] The main purpose of CAD is to increase the accuracy of diagnosis as well as improve the consistency of image interpretation by using the computer results as a "second opinion". Here a major challenge in employing image mining techniques is to effectively relate low-level features (automatically extracted from image pixels) to high-level semantics based on the human perception. This is a research area that has received considerable attention in recent years.[12,17]

In the medical imaging domain, digital mammograms are among the most difficult images to be read due to their low contrast and differences in the types of tissues.[18] A vast number of mammograms is generated daily in hospitals and medical centers. Thus, the radiologists have more and more images for manual analysis. After analyzing a number of images, the process of diagnosis becomes tiresome and consequently becomes more susceptible to errors. On the other hand, important visual clues of breast cancer, including preliminary signs of masses and calcification clusters, are very subtle in the early stages of breast cancer, making diagnosis difficult and challenging even for experts. This is the main reason for the development of automated classification systems to assist medical staff. Within the literature, much research has been focused on automated classification methods, but with all this effort, there is still no widely used method to classify mammograms. This is due to the fact that the medical domain requires high accuracy, and especially, the rate of false negatives has to be very low. In the case of mammography, the technique alone cannot prove that a suspicious area is malign or benign.[17] To determine whether the suspicious area is malignant, the tissue has to be removed for examination using breast biopsy techniques. A false-positive detection may cause an unnecessary biopsy. Statistics show that only 20−30% of breast biopsy cases are proved cancerous. Fine needle aspiration cytology (FNAC) is also widely adopted in the diagnosis of breast cancer, but the average correct identification rate of FNAC is only 90%.[11] In a false-negative detection, an actual tumor that remains undetected could lead to higher costs or even cost a human life. All these reasons make the decisions to be made on mammograms difficult.

Automated classification of mammograms has gradually attracted considerable attention from computer science practitioners and researchers, especially by those employing association rules.[9,10,15–17,19,20] However, no study has gone through the separate mining of image blocks, which in turn could improve the performance of the classifier, in particular by reducing the number of false-negative detections. In the present study, an associative classification framework based on parallel mining of non-overlapping image blocks is proposed to be used for mammograms discrimination. Indeed, association rules mining is applied to a commonly used mammography image database to classify digital mammograms into three categories, namely normal, benign and malign. In order to do so, images are first preprocessed and then the features are extracted from non-overlapping image blocks and discretized for rule discovery. Association rules are then discovered through parallel mining of transactional databases, which correspond to the image blocks, and finally are used within a unique decision-making scheme to predict the class of unknown samples. The experimental results show that the proposed framework proved successful in mammograms discrimination, incurring an average accuracy as high as 85%, along with average precision and recall of 88.80% and 90.20%, respectively.

Background and Related Work

Breast cancer is one of the most common cancers, causing death among women, especially in developed countries.[21] This type of cancer has the second highest rate of death among cancers in women, and the chance of a woman having invasive breast cancer some time during her life is about one in eight. It is also said that the chance of dying from breast cancer is about one in 35.[20,22] Women with breast cancer in their family are more susceptible to developing breast cancer. The risk also increases with age, not having children, and obesity.[11,20] Though much less common, breast cancer also occurs in men.[23] Mammography, as a medical imaging modality, can help the early detection of breast cancer, thereby increasing patient survival rates. In some cases, mammography is the first step in preventing the spread of cancer through early detection, and in many cases makes it possible to cure or eliminate the cancer altogether. On the other hand, digital image processing provides the possibility of efficient storage, easy image retrieval and the maintenance of large image databases for both educational and research purposes.[16] This together has provided an opportunity to employ image mining techniques, such as classification, for purposes including the support and development of CAD systems.

Automated classification of mammograms has gradually emerged as an appealing domain for the evaluation of design and implementation of CAD systems, and this includes studies based on association rules. For example, previously a new breast mass classification method based on quantitative association rule mining has been developed.[19] In another attempt, an associative classification algorithm to generate atomic rules under high, self-adaptive confidence thresholds and dynamic support thresholds was developed, where an image was segmented into $n \times n$ regions, and 19 features were extracted from these segments.[9] Also, multilevel association rules has been already applied to hierarchically clustered objects from various images, performing object based segmentation on the images.[10] Recently, association rules have attracted interest as a means to achieve multiclass discrimination of mammograms. In this respect, a method has been presented to generate suggestions for diagnosis of mammograms, which could be integrated into a CAD system.[16,20,24] In their work, the authors combined low-level features automatically extracted from images and high-level knowledge from experts to search for patterns. Their approach was divided into four main steps: feature extraction, discretization and feature selection, association rule mining, and generation of diagnosis suggestions. Also, in another work, a method for the classification of mammograms using a unique weighted association-rule-based classifier was introduced.[12] The method followed many image processing techniques applied to images, and then texture components were extracted from image segments. The extracted continuous features were then discretized over 10 fixed equal-width intervals for rule discovery, and the association rules were adopted for classification.

Problem Statement

A drawback of the works reviewed above is that, after images are segmented into regions, association mining is applied to a transactional database built from features extracted from the segments all together, rather than mining the image segments separately. Indeed, mapping the whole mammogram into a single high-dimensional feature space, hence a single transactional database to be mined, results in an undesirable performance of the associative classifier in favor of the normal class. Let X be an abnormal mammogram, segmented into four non-overlapping blocks, to be categorized by an associative classifier. In most of the abnormal mammograms, the area of abnormality is presented in a region within one of the four image blocks. As a result, in our case, there would be one image block with the characteristic of abnormal tissue, whereas the other three image blocks would be normal. Therefore, X when modeled by a set of features as a new transactional itemset, triggers a greater number of associative classification rules whose consequents refer to the normal class, despite X being abnormal. This leads into a false-negative detection, which in turn suppresses the precision of the classifier in the case of normal samples, something that is undesirable in medical decision-making. To circumvent the problem, in the present study, an associative classification framework is proposed to be used for mammograms discrimination based on parallel mining of non-overlapping image blocks. Experiments are done on a widely used mammography image database to evaluate the effectiveness of the classifier in terms of accuracy, precision and recall. Results show that the proposed framework proved successful in discriminating mammograms.

MATERIALS AND METHODS

This section introduces the building blocks of the methodology undertaken in the present study. The aim is to propose a novel associative classification framework which can discriminate mammograms effectively, in particular by reducing the number of false negative detections. As it is shown in Fig. 1, our method is composed of five stages, namely *image preprocessing*, *feature extraction*, *data transformation*, *parallel mining of image blocks* and finally *associative classification of unknown samples*. Indeed, images are first modeled as a set of transactions, each transaction representing one image with the features along with the class label. Association rules are then discovered through parallel mining of transactional databases which correspond to image blocks, and the results are used within a unique decision-making scheme to predict the class of unknown samples.

The Dataset

The data collection used in our experiments was taken from the mammography image analysis society (MIAS) database, which is commonly used for mammograms classification.[25] There are a total of 330 images from three classes, namely: normal, benign and malign, available in the database. Within the images, 64% belong to normal class, 20% to benign, and 16% are those of malign class. The abnormal cases (i.e. benign and malign ones) are further divided into six categories: microcalcification, circumscribed masses, spiculated masses, ill-defined masses, architectural distortion and asymmetry. All the images also include the locations of any abnormalities that may be present. Furthermore, data in the database consists of the location of the abnormality (like the center of a circle surrounding the tumor), its radius, breast position (left or right), type of breast tissues (fatty, fatty-glandular and dense) and tumor type, if it exists (benign or malign). All the mammograms are taken from medio-lateral oblique view.[17]

Image Preprocessing

This stage corresponds to data preparation, which is usually involved in practical data mining problems. Since real-life raw data is often noisy, containing deliberately inserted identifiable labels, preprocessing becomes a necessity prior to feature extraction and rule discovery. Such artifacts can often lead to redundant and non-informative rules. Furthermore, data of type medical images are usually large in terms of pixels, and most of the image plane consists of homogeneously colored background, which does not provide any diagnostic information. In our case, we had images in the size of 1024×1024, with partial noise and unwanted background. In addition, these images were scanned at different illumination conditions, and therefore some images appeared too bright and some were too dark. Hence, the first step toward noise removal was to prune

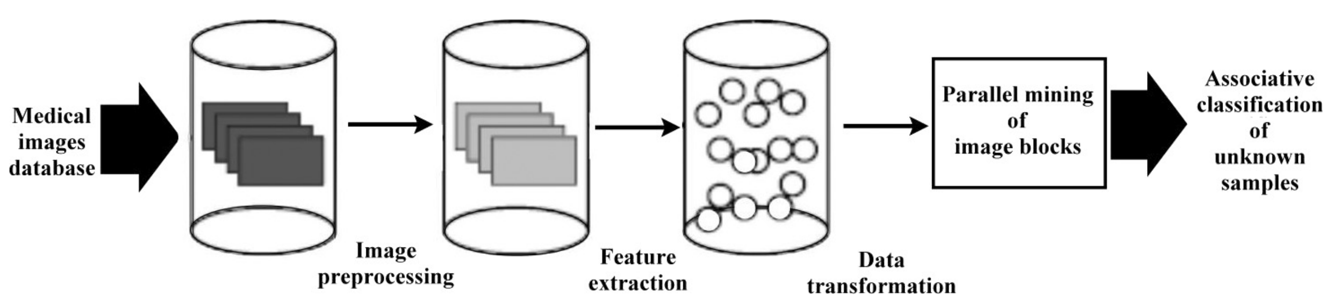
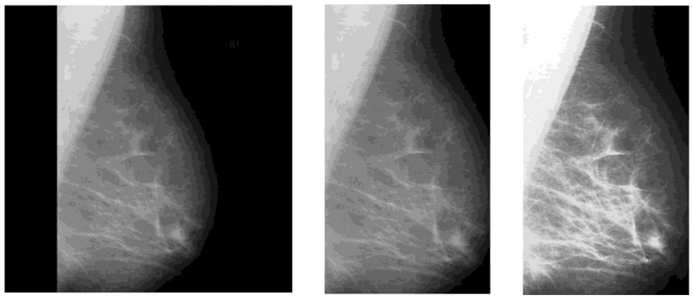


Fig. 1 Block diagram of the proposed framework.



(A) **(B)** **(C)**

Fig. 2 Preprocessing a given right mammogram: **(A)** raw image; **(B)** cropped image; **(C)** enhanced image.

the images by means of the Crop operation, which is a commonly used tool in image processing. Cropping cuts the unwanted portions of the image. Thus, we eliminated almost all the background information and most of the noise. The next step to complete the preprocessing stage was to enhance the cropped images, which was done in the spatial domain. Image enhancement helps in qualitative improvement of the image with respect to a specific application.[17] We employed the Histogram Equalization technique to highlight the features to be extracted from image blocks. Figure 2 shows the result of image preprocessing for a given mammogram.

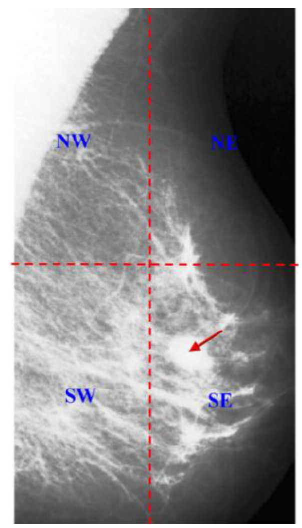


Fig. 3 Segmentation of a given right mammogram into four non-overlapping blocks prior to feature extraction.

Feature Extraction

In the context of data mining, each individual independent instance is characterized by its values on a fixed, predefined set of features or attributes, which together constitute the input to the learning method. There are features whose values are numeric or continuous and those that are nominal or categorical. In the present study, we first segmented each preprocessed image as a sample into four non-overlapping blocks, referred to as *North West* (*NW*), *North East* (*NE*), *South West* (*SW*) and *South East* (*SE*), using equal-width approach, as shown in Fig. 3. This was done in order to scope out the mammographic regions, something that would also resemble the intuitive association of image blocks to target classes which usually happens during the examination of mammograms by an expert. Then, for the normal images, features of all the four segmented blocks were extracted, whereas for those characterizing an abnormal mammogram, only features of the abnormal image block were extracted (e.g. for the mammogram presented in Fig. 3 only the features extracted from the *SE* block (the arrow in the region points to the tumor) were extracted). The extracted features are four statistical image parameters: Mean, Variance, Skewness and Kurtosis, introduced in Table 1. Thereby, each image block represents a feature vector of the length of four. After the *data transformation* stage, the feature vectors are used in place of the training samples to model them as transactions.

Table 1. Texture Features Used to Represent Image Blocks.

Feature Label	Features	Formula
F_1	Mean	$M = \frac{\sum x}{n}$
F_2	Variance	$\text{Var} = \frac{\sum (x-M)^2}{n}$
F_3	Skewness	$\text{Skew} = \frac{1}{n} \times \left(\frac{\sum(x-M)}{\sigma}\right)^3$
F_3	Kurtosis	$\text{Kurt} = \frac{1}{n} \times \left(\frac{\sum(x-M)}{\sigma}\right)^4 - 3$

Note: Here x is a pixel value, n is the number of pixels, and σ is the standard deviation.

Data Transformation

Once the features are extracted, we need to transform the input data so that it can be used for association rules mining. Data transformation (engineering the input data into a form suitable for the learning method) is commonplace in most of the data mining and machine learning tasks, before applying a particular technique to the available data.[6] This includes, for example, normalization, outlier detection, feature selection or discretization.[1] With respect to the latter, some classification and clustering techniques deal with nominal attributes only and cannot handle ones measured on a numeric scale. This applies, for instance, to decision tree and association rules learners. To use them on general datasets, numeric attributes must first be "discretized" into a small number of distinct ranges. Even methods that can handle numeric attributes often produce better results, or work faster, if the attributes are prediscretized. After discretization, interval labels can be used to replace actual data values. Replacing numerous values of a continuous attribute by a small number of interval labels thereby reduces and simplifies the original data. This leads to a concise, easy-to-use, knowledge-level representation of mining results. But, here the question arises: what is a good way to discretize numeric attributes into ranges before any learning takes place?

When the domain of values in a set of quantitative attributes is large, an obvious approach will be to first partition the values into intervals and then map each (*attribute*, *interval*) pair to a Boolean attribute.[26] We can now use any algorithm for finding Boolean association rules to find quantitative association rules. However, there are two problems with this simple approach when applied to quantitative attributes: if the intervals are too large, some rules may not have minimum confidence, which leads to information loss; and if they are too small, some rules may not have minimum support. In other words, there is a trade-off between faster execution time with fewer intervals, and reducing information loss with more intervals. We can reduce the information loss by increasing the number of intervals at the cost of increasing the execution time and potentially generating many uninteresting rules. The simplest and most used discretization technique divides the range of observed attribute values into k equal-sized intervals. In this case, in a previous study, the optimal number of intervals has been established, given an arbitrary measure of the maximum information loss due to the discretization, the so-called *partial completeness*.[25] In that study, it is demonstrated that equal-width partitioning (also known as equal-interval binning) is, in some sense, optimal for this measure of partial completeness.

In the present study, the continuous feature values were discretized into k equal-sized intervals, and this was done for different values of k. Indeed, k was treated as a free parameter of the association learner in order to find the best choice of value for the parameter which would yield the best result. The training samples for each feature f_i over all the classes were combined to find the minimum ($\min\_f_i$) and maximum ($\max\_f_i$) values of f_i. These minimum and maximum values then were used to discretize the continuous values of f_i into k equal-width intervals. The width of each interval was calculated using the formula $l_i = (\max\_f_i - \min\_f_i)/k$. Next, a two-field substitute code was used to represent values in each bin by means of a coding paradigm (Fig. 4). Accordingly, item "1403", for example, represents the third interval of feature VarSE. As a result, continuous attributes were discretized into intervals, and intervals in turn were mapped to a numerical code as an item.

Finally, the resultant itemsets were organized in the form of transactions, which in turn constitute the input for the association rules mining algorithm. The transactions are of the form $\{RF_1, RF_2, RF_3, RF_4, ClassLabel\}$, where RF_i is a discretized regional feature extracted from an image block, and $ClassLabel$ is the discrimination of the image block as normal, benign or malign. The following are samples of transactional itemsets:

103	204	308	405	1
102	201	305	406	10
101	202	308	408	100
505	605	708	802	1
506	605	707	801	10
506	603	707	802	100
904	1005	1107	1201	1
905	1009	1106	1201	10

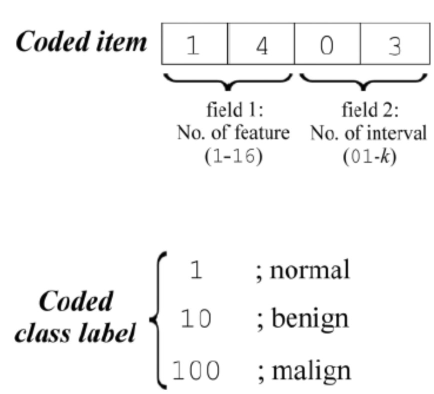


Fig. 4 Coding paradigm used to obtain itemsets.

908	1006	1104	1202	100
1307	1404	1506	1602	1
1307	1406	1503	1604	10
1305	1406	1510	1604	100

Parallel Mining of Image Blocks

Association rules mining aims at finding interesting relationships in the form of associative rules among sets of data items, and in a probabilistic sense, shows items that occur frequently together in a given dataset. The problem of mining association rules is formally stated as follows. Let $I = \{i_1, i_2, \ldots, i_m\}$ be a set of items. Let T be a set of transactions, where each transaction t is a set of items such that $t \subset I$. A transaction t contains X, a set of some items in I, if $X \subset I$. An association rule is an implication of the form $X \Rightarrow Y$ where $X \subset I$, $Y \subset I$, and $X \cap Y = \emptyset$. The rule $X \Rightarrow Y$ holds in T with confidence C if the fraction of transactions that also contain Y in those which contain X in T is C. The rule $X \Rightarrow Y$ has support S in T if the fraction of transactions in T that contain $X \cup Y$ is S. Given a set of transactions T, the problem of mining association rules is to generate all association rules that have support and confidence no less than the user-specified *minimum support* and *minimum confidence*, respectively. Once frequent itemsets are obtained, it is straightforward to generate association rules with confidence no less than *minimum confidence*. Hence, association rules mining is normally a two-step process, where in the first step frequent itemsets are discovered (i.e. itemsets whose support is no less than *minimum support*) and in the second step association rules are derived from the frequent itemsets. The *a priori* algorithm, first introduced by Agrawal *et al.*,[3] is the most common algorithm that is designed to implement association rules mining in a large database. The *a priori* algorithm works iteratively. It first finds the set of large one-itemsets, and then set of two-itemsets, and so on.[3]

In many applications, the primary problem in using association rules mining concerns using global minimum support. Rare items will not be included in the frequent sets because it will not hold enough support. One solution is to have a very small support threshold; however, we will end up with a very large number of frequent itemsets, which is computationally hard to handle. Therefore, specifying multiple thresholds may allow rare transactions, which might be very important, to be included in the frequent itemsets. Other issues might arise, depending on the application itself.[27] In the case of mining mammography image database, for example, two data organizations has been proposed, which could be used during the mining process.[19] The first one, called association rule-based classification with all categories (ARB-AC), is applied when the rules are extracted from the entire training set at once. In this approach, all the transactions in the database form a single training set, and the rules generated are *de facto* the classier. In the second organization, on the other hand, mining is done based on the data classes instead of the entire dataset at once i.e. each class is considered as a separate training set and the mining algorithm is applied to it. So, this data organization is referred to as association rule-based classification by category (ARB-BC). However, the first organization, when applied to a medical image database such as Mammography Image Analysis Society (MIAS), suffers from the very small relative number of abnormal training samples in contrast with the normals, since normal observations always occur more frequently in the real world. Therefore, no or very small number of frequent patterns belonging to abnormal classes would be discovered as the result of mining unbalanced data. Also, it has been reported that in the case of the mammography image database, the classification accuracy reached when ARC-BC is used is higher than the one obtained when the training set was mined at once using ARC-AC, which shows the importance of data organization in the case of mining medical images.[17]

In our approach, we used the *a priori* algorithm in order to discover association rules among the data items extracted from the mammograms and the class to which each mammogram belongs. In order to do so, we first organized the itemsets as the modular transactional databases depicted in Fig. 5. In this new type of data organization, we took advantage of ARB-BC association mining. As it is shown in Fig. 5, classification rules for image blocks are discovered through parallel mining of their corresponding transactional databases. During this mining process, the association rules were explored by setting the *minimum support* at 25% for abnormal transactions and 10% in the case of normal transactions. Since association mining was applied by class categories separately, confidence for the discovered rules would be 100%. Also we constrained the association rules such that the antecedent of the rules is composed of a conjunction of data items from the mammogram, while the consequent of the rule is always the class to which the mammogram belongs. In other words, a rule would describe a frequent set of data items per class based on the *a priori* algorithm. The following are some sample discovered rules:

306 $\rightarrow$1 Sup=26.59% Conf=100%
406 $\rightarrow$10 Sup=40% Conf=100%
306 $\rightarrow$100 Sup=38.47% Conf=100%

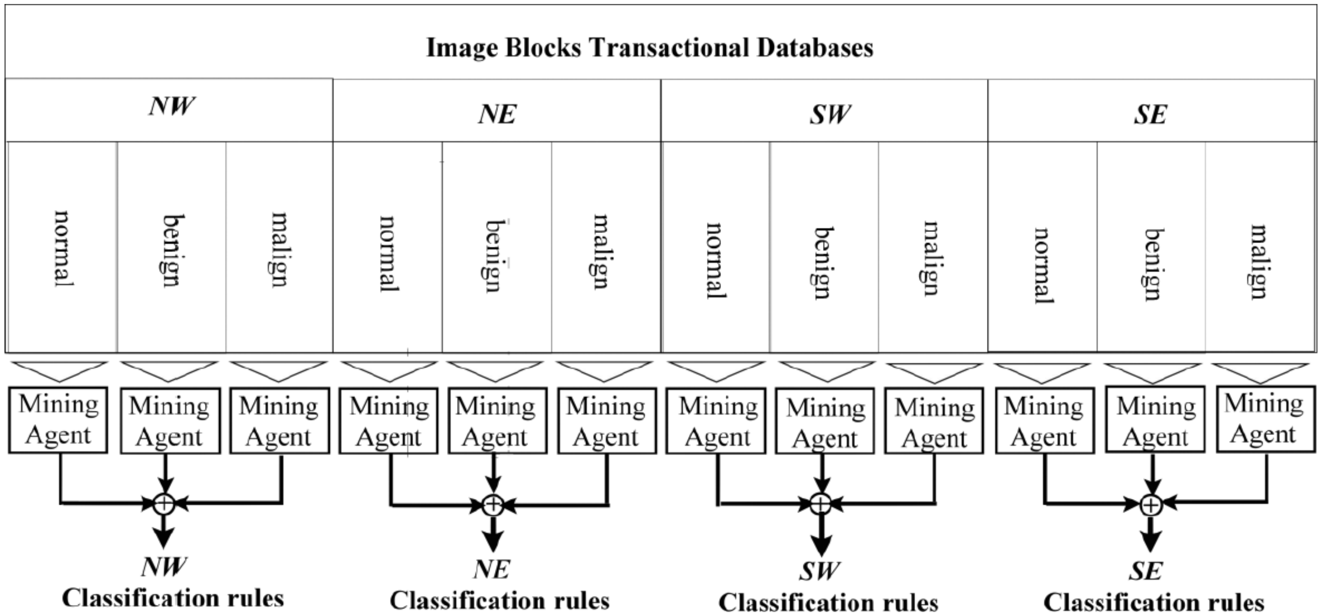


Fig. 5 Parallel mining of transactional databases obtained from image blocks.

802 →1 Sup=45.57% Conf=100%
707 801 →10 Sup=33.34% Conf=100%
604 802 →100 Sup=33.34% Conf=100%
906 1107 1201 →1 Sup=11.40% Conf=100%
1107 1201 →10 Sup=40% Conf=100%
1110 1203 →100 Sup=28.58% Conf=100%
1306 1506 →1 Sup=11.40% Conf=100%
1506 1605 →100 Sup=28.58% Conf=100%

Our framework was applied once to the left mammograms and once to the right mammograms in each training session, and the classification rules obtained each time were used to predict the class of test samples.

RESULTS

Associative Classification of Unknown Samples

In this section, classification of test samples is described, where association rules generated and stored previously are analyzed within a decision-making scheme to make the final predictions. The general idea behind the decision-making scheme is to search for the class of a new image in a sequence which is governed by the association rules obtained for each image block, as it is shown in Fig. 6.

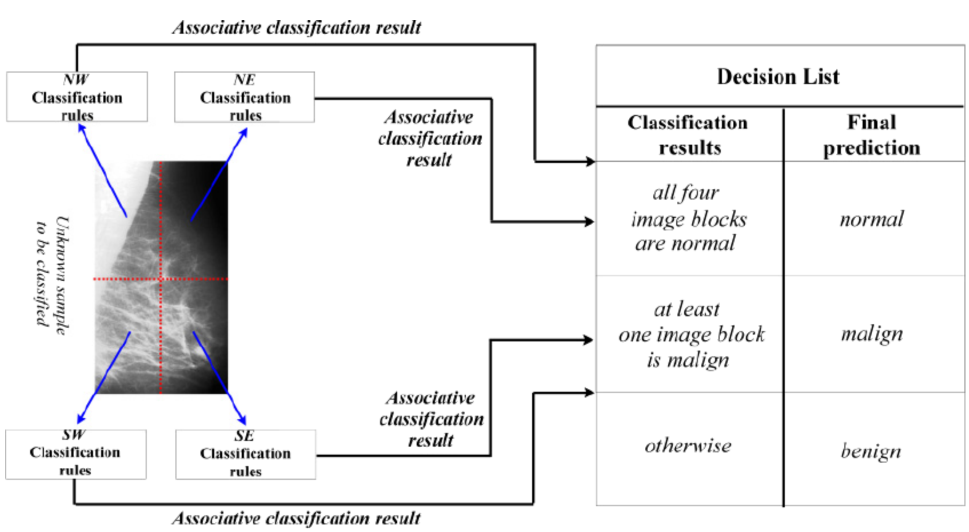


Fig. 6 Decision-making scheme employed to predict the class of unknown samples.

Association classification is a special case of association rule mining in which only the class attribute is considered in the rule's consequent. Methods of associative classification differ primarily in the approach used for frequent itemset mining, and in how the derived rules are analyzed and used for classification.[1] There are various methods for associative classification, including classification-based association (CBA) and classification based on multiple association rules (CMAR). However, a solution for classifying a new transaction is to attach the class that has the most rules matching this new transaction, or the class associated with the first high ranked rule that applies to the new sample.[17] In our approach, we used the former strategy to predict the class of unknown samples. Suppose that we are given an image X to be classified by the system (Fig. 6). First features are extracted and a transaction of data items are created as discussed before. Then, each itemset representing one of the four image blocks is sent to the corresponding set of classification rules to be labeled as normal, benign or malign. Here, the classification process counts the number of rules that satisfy or match the itemset, and then assigns a label according to the class with the most number of matched rules. Finally, the classification results are included in a decision list and used to make the final prediction. If all the image blocks were categorized as normal, then the image X is classified as normal. However, if at least one of the image blocks was recognized as malign, the image would be classified as malign. Otherwise, X is predicted to be benign. In cases where no rule matches the test image condition or an equal number of rules are fired, the classification process, and hence the prediction, result in inconclusive recognition. Therefore, in the fashion described above, the learned classification rules are used within a decision making scheme which together construct the classification model.

System Evaluation

In applications such as medical images classification, the real problem occurs when there is no vast supply of data available to be used for system evaluation. Therefore, the problem becomes how to make use of a dataset with a limited number of observations in its best way. One way to tackle the problem is that a certain amount of data is held over for testing — so-called *holdout* procedure — and the remainder is used for training (and, if necessary, part of that is set aside for validation).[6] In this case, the estimate of the classifier accuracy is pessimistic because only a portion of the initial data is used to derive the model. Hence, random subsampling is usually adopted as variation of the holdout method, in which the holdout method is repeated n times. The overall accuracy estimate is then taken as the average of the accuracies obtained from iterations. In our study, we employed the same technique to assess the accuracy of the proposed associative classifier. We divided the dataset into five partitions to perform the experiments, once on the left mammograms and once on the right mammograms. For each partition, we selected about 86% of the dataset for training and the rest for testing. During each training/testing session, a fixed value was chosen for the associative learner parameter k. This was done by starting from four and continuing such that the value was increased by one for the next session, until no rules were extracted for a target class. Finally, the results for the best case were reported. Table 2 shows the results obtained using our framework for each dataset partition, along with the value of k for which the best recognition rate was achieved and the number of extracted rules from image blocks. Based on the results, the recognition rate for the associative classifier was 85.06% in the case of right mammograms and 79.67% for the left mammograms on average.

We also present precision and recall graphs in Fig. 7 to show that both false-positive and false-negative detections are low for normal cases, which is very desirable in medical decision-making. The values for precision and recall are given by:

$$\text{Precision} = \frac{TP}{TP + FP}$$

and

$$\text{Recall} = \frac{TP}{TP + FN}$$

In this case, TP = images which are normal and are labeled normal by the classifier; FP = images which are abnormal, but are labeled normal; and FN = images which are normal, but are labeled abnormal. Figure 7 shows the obtained values for precision and recall over the five dataset partitions for both left and right mammograms. From Fig. 7, it can be seen that the values are fairly high for all training/testing sessions.

Results Interpretation

This section aims to look at the results obtained in our study and to accentuate the primary factors which play significant roles in our proposed framework, and hence deserve more consideration when the problem is applying associative classification to medical images. First, the results presented in Table 2 illustrate the important

Table 2. Results Obtained by Applying the Proposed Framework to Different Dataset Partitions in the Case of: (A) Left Mammograms, (B) Right Mammograms.

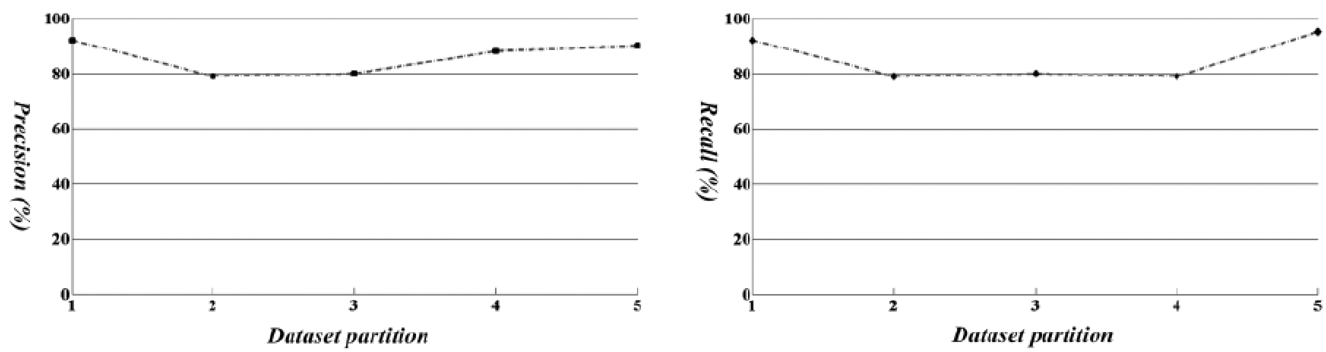
(A)

			No. of Rules Extracted from Image Blocks			
Dataset Partition	***k***	**Recognition Rate%**	**NW**	**NE**	**SW**	**SE**
1	12	84.62	31	29	28	30
2	6	65.22	52	55	47	56
3	12	67	29	29	28	30
4	10	62	37	39	28	33
5	11	89.48	32	26	34	30

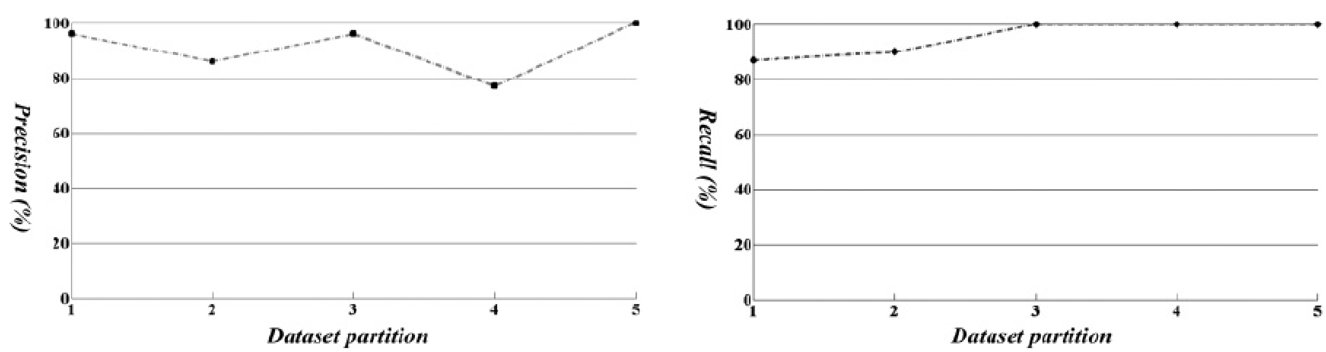
(B)

			No. of Rules Extracted from Image Blocks			
Dataset Partition	***k***	**Recognition Rate%**	**NW**	**NE**	**SW**	**SE**
1	10	84.64	23	34	56	28
2	9	78.27	23	34	53	34
3	9	91.31	25	34	51	32
4	9	76.93	25	42	57	34
5	9	94.45	26	34	52	35

role of discretization in developing associative image classifiers which employ continuous features, and to be exact, the suitable choice of value for discretization parameter k. Also one can infer from the results that the number of extracted rules depends on the value chosen for this parameter. It was observed during the experiments that the number of mined rules generally tends to decrease as the discretization becomes narrower by selecting a greater value for k. Second, we noticed in our study that the choice of appropriate value for the support threshold can also have a significant effect on the accuracy of the associative classifier. In our case, we concluded to the values of 10% for mining normal image blocks and 25% for abnormal ones. However, it might be



(A)



(B)

Fig. 7 Precision and recall graphs for: **(A)** left mammograms; **(B)** right mammograms.

possible to devise an algorithm to help us to find an optimum threshold value to be used during the mining process.[7] The two primary factors introduced earlier demonstrate that successful data mining involves far more than selecting a particular technique and applying it to the available data. Many learning methods have various parameters for which suitable values must be chosen.[6] In most cases, results can be improved markedly by suitable choice of parameter values, and the appropriate choice depends on the data at hand.

Finally, it was revealed during our experiments that proper data organization, block-based mining of images and decision-making scheme by which the final predictions are made could improve the performance of the classifier, in particular with respect to precision and recall. The decision-making scheme could be chosen from a range of decision-making strategies, but it should have a high level of meaningfulness from the medical point of view. In our study, we managed to predict the class of unknown samples through parallel mining of image blocks and providing a scheme that models the intuitive associations which usually happen during examination of mammograms by an expert. Together, they had a significant effect on improving the performance of the classifier. However, extensions are needed in order to approach to a more concrete model which would yield better results. The next section summarizes some of the future possibilities that could refine the proposed associative classification framework.

CONCLUSION

A vast amount of mammograms generated in hospitals and medical centers calls for developing CAD systems to assist the medical staff. This is due to the fact that mammograms classification is a difficult diagnostic task particularly when used in screening programs. Within the literature, much research has been focused on automated classification of mammograms, but with all this effort, there is still no widely used method. This is because medical domain requires high accuracy and, especially, the rate of false negatives to be very low. In this paper, an associative classification framework based on parallel mining of non-overlapping image blocks was proposed to be used for mammograms discrimination. The experimental results show that the proposed framework proved successful in terms of accuracy, precision and recall. However, extensions are needed in order to approach to a more concrete model which would yield better results. Indeed, the proposed framework in this paper is still not able to deal with the data transformation effectively, which uses a fixed value for the discretization parameter. It is worth pointing out that discretized intervals of the same size can imply the loss of significant information to mine rules with high support and confidence. Equal-interval binning often distributes instances very unevenly: some bins contain many instances, and others contain none. This can seriously impair the ability of the attribute to help building good decision structures. Hence, future work includes developing an algorithm to handle the discretization process in its best way. Further studies can also be made with respect to the decision-making scheme to include an incremental modification to the rule extraction or firing. Indeed, the classification rules require modification so that the associative classifier could be adjusted in the case of misclassifications. This is mostly needed to tackle the uncertainty involved in the classification task. One reason that makes uncertainty appear in the prediction is imperfections which exist in raw input data needed for knowledge discovery, something that is inherent in medical images. One way to confront the problem could be introducing some hybridization of the proposed framework with a soft computing method. Additionally, extraction of more meaningful textural features could help in the discovery of more specific rules in terms of between-class separability.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support provided by Science and Research Branch of Islamic Azad University, Tehran, Iran, to carry out the research.

REFERENCES

1. Han J, Kamber M, *Data Mining: Concepts and Techniques*, 2nd edn., Morgan Kaufmann Publishers, Elsevier Inc, USA, 2006.
2. Kantardzic M, *Data Mining: Concepts, Models, Methods, and Algorithms* (John Wiley and Sons Inc, USA, 2003).
3. Agrawal R, Imielinski T, Swami A, Mining association rules between sets of items in large databases, *Proc ACM SIGMOD ICMD*, pp. 207–216, 1993.
4. Chen G, Wei Q, Liu D *et al.*, Simple association rules (SAR) and the SAR-based rule discovery, *Comput Ind Eng* **43**:721–733, 2002.
5. Chen YL, Tzu-Hsuan Hung L, Using decision trees to summarize associative classification rules, *Expert Syst Appl* **36**:2338–2351, 2009.
6. Witten IH, Frank E, *Data Mining: Practical Machine Learning Tools and Techniques* (Morgan Kaufmann Publishers, USA, 2005).

7. Coenen F, Leng P, The effect of threshold values on association rule based classification accuracy, *Data Knowl Eng* **60**:345−360, 2007.
8. Mitra S, Acharya T, *Data Mining: Multimedia, Soft Computing, and Bioinformatics* (John Wiley and Sons Inc, USA, 2003).
9. Xu X, Han G, Huaqing M, A novel algorithm for associative classification of image blocks, *Proc 4th Int Conf Comput Inf Tech*, pp. 46−51, 2004.
10. Tseng VS, Wang MH, Su JH, A new method for image classification by using multilevel association rules, *Proc 21st Int Conf Data Engineering*, p. 180, 2005.
11. Karabatak M, Inceb MC, Sengur A, Wavelet domain association rules for efficient texture classification, *Appl Soft Comput* **11**:32−38, 2009.
12. Dua S, Singh H, Thompson HW, Associative classification of mammograms using weighted rules, *Expert Syst Appl* **36**:9250−9259, 2009.
13. Rajendran P, Madheswaran M, Pruned associative classification technique for the medical image diagnosis system, *Proc 2nd Int Conf Machine Vision*, pp. 293−297, 2009.
14. Rajendran P, Madheswaran M, Naganandhini K, An improved pre-processing technique with image mining approach for the medical image classification, *Proc 2nd Int Conf Comput Commun Network Technol*, pp. 1−7, 2010.
15. Dua S, Jain V, Thompson HW, Patient classification using association mining of clinical images, *Proc 5th IEEE Int Symposium on Biomed Imaging: From Nano to Macro*, pp. 253−256, 2008.
16. Ribeiro MX, Bugatti PH, Traina Jr C. *et al.*, Supporting content-based image retrieval and computer-aided diagnosis systems with association rule-based techniques, *Data Knowl Eng* **68**:1370−1382, 2009.
17. Antonie ML, Zaiane OR, Coman A, Associative classifiers for medical images. LecNICS 2797, MMCD, pp. 68−83, Berlin/Heidelberg: Springer, 2003.
18. Zaiane OR, Antonie ML, Coman A, Mammography classification by an association rule-based classifier, International Workshop on Multimedia Data Mining (with ACM SIGKDD), pp. 62−69, 2002.
19. Wang X, Smith MR, Rangayyan RM, Mammographic information analysis through association-rule mining, Canadian Conference on Electrical and Computer Engineering-CCGEI, pp. 1495−1498, Niagara Falls, May 2004.
20. Ribeiro MX, Traina AJM, Traina Jr C. *et al.*, An association rule-based method to support medical image diagnosis with efficiency, *IEEE Trans Multimedia* **10**(2):277−285, 2008.
21. Roseline R, Thangavel K, Classification ensemble for mammograms using ant-miner, *Proc 2nd Int Conf Comput Commun Network Technol*, pp. 1−6, 2010.
22. Senthilkumar J, Kavitha JK, Manjula D *et al.*, ADMID: An association rule discovery for mammogram image diagnosis, *Proc 22nd IEEE Int Symposium on Comput-based Med Syst* pp. 1−8, 2009.
23. Antonie ML, Zaiane OR, Coman A, Application of data mining techniques for medical image classification, *Proc 2nd Int Workshop on Multimedia Data Mining in Conjunction with ACM SIGKDD Conf* San Francisco, USA, pp. 94−101, 2001.
24. Ribeiro MX, Traina AJM, Balan AGR *et al.*, Sugar: A framework to support mamm ogram diagnosis, *Proc 12th IEEE Int Symposium Comput-Based Med Syst*, pp. 47−52, 2007.
25. Mammography Image Analysis Society (MIAS). Available at: http://peipa.essex.ac.uk/info/mias.htm, March 2012.
26. Srikant R, Agrawal R, Mining quantitative association rules in large relational tables, *SIGMOD* Montreal, Canada, pp. 1−12, 1996.
27. Awad M, Khan L, Thuraisingham B *et al.*, *Design and Implementation of Data Mining Tools* Taylor and Francis, USA, 2009.