Associative Classifiers for Medical Images *

Maria-Luiza Antonie, Osmar R. Zaïane, and Alexandru Coman

Department of Computing Science
University of Alberta
Edmonton, Alberta, Canada
{luiza, zaiane, acoman}@cs.ualberta.ca

Abstract. This paper presents two classification systems for medical images based on association rule mining. The system we propose consists of: a pre-processing phase, a phase for mining the resulted transactional database, and a final phase to organize the resulted association rules in a classification model. The experimental results show that the method performs well, reaching over 80% in accuracy. Moreover, this paper illustrates how important the data cleaning phase is in building an accurate data mining architecture for image classification.

1 Introduction

Association rule mining is one of the most important tasks in Data Mining and it has been extensively studied and applied to market basket analysis. In addition, building computer-aided systems to assist medical staff in medical care facilities is becoming of high importance and priority for many researchers. This paper describes the use of association rule mining in an automatic medical image classification process.

This paper presents two classification methods for medical images. It is based on association rule mining and it is tested on real datasets in an application for classifying medical images. This work is a significant extension and improvement of the system and algorithm we developed and presented in [2]. The novelty is in the data cleaning and data transformation techniques as well as in the algorithm used to discover the association rules. This work illustrates the importance of data cleaning in applying data mining techniques in the context of image content mining.

The high incidence of breast cancer in women, especially from developed countries, has increased significantly in recent years. The etiologies of this disease are not clear and neither are the reasons for the increased number of cases. Currently there are no methods to prevent breast cancer, that is why early detection represents a very important factor in cancer treatment and allows reaching a high survival rate. Mammography is considered the most reliable method in early detection of cancer. Due to the high volume of mammograms

* Maria-Luiza Antonie was partially supported by Alberta Ingenuity Fund and Osmar R. Zaïane was partially funded by NSERC, Canada

to be read by physicians, the accuracy rate tends to decrease and automatic reading of digital mammograms becomes highly desirable. It has been proven that double reading of mammograms (consecutive reading by two physicians or radiologists) increased the accuracy, but at high costs. That is why the computer aided diagnosis systems are necessary to assist medical staff to achieve high efficiency and effectiveness.

The methods proposed in this paper classify the digital mammograms into three categories: normal, benign and malign. The normal ones are those characterizing a healthy patient, the benign ones represent mammograms showing a tumor, but that tumor is not formed by cancerous cells, and the malign ones are those mammograms taken from patients with cancerous tumors. Generally, the most errors occur when a radiologist must decide between the benign and malign tumors. Mammography reading alone cannot prove that a suspicious area is malignant or benign. To decide, the tissue has to be removed for examination using breast biopsy techniques. A false positive detection causes an unnecessary biopsy. Statistics show that only 20-30% of breast biopsy cases are proven cancerous. In a false negative detection, an actual tumor remains undetected that could lead to higher costs or even to the cost of a patient's life.

Digital mammograms are among the most difficult medical images to be read due to their low contrast and differences in the types of tissues. Important visual clues of breast cancer include preliminary signs of masses and calcification clusters. Unfortunately, at the early stages of breast cancer, these signs are very subtle and varied in appearance, making diagnosis difficult, challenging even for specialists. This is the main reason for the development of classification systems to assist specialists in medical institutions. Since the data that physicians and radiologists must deal with has increased significantly, there has been a great deal of research done in the field of medical images classification. With all this effort, there is still no widely used method to classify medical images. This is because this medical domain requires high accuracy. Also, mis-classifications could have different consequences. False negatives could lead to death while false positives have a high cost and could cause detrimental effects on patients. For automatic medical image classification, the rate of false negatives has to be very low if not zero. It is important to mention that manual classification of medical images by professionals is also prone to errors and the accuracy is far from perfect. Another important factor that influences the success of automatic classification methods is working in a team with medical specialists, which is desirable but often not achievable. The consequences of errors in detection or classification are costly.

In addition, the existing tumors are of different types. These tumors are of different shapes and some of them have the characteristics of normal tissue. All these things contribute to the decisions that are made on such images even more difficult. Different methods have been used to classify and detect anomalies in medical images, such as wavelets [4, 15], fractal theory [8], statistical methods [6] and most of them used features extracted using image processing techniques [13]. In addition, some other methods were presented in the literature based on fuzzy set theory [3], Markov models [7] and neural networks [5, 9]. Most of the computer-aided methods proved to be powerful tools that could assist medical staff in hospitals and lead to better results in diagnosing a patient.

The remainder of the paper is organized as follows. Section 2 describes the feature extraction phase as well as the cleaning phase. The following section presents the new association rule-based method used to build the classification system. Section 4 describes how the classification system is built using the association rules mined. Section 5 introduces the data collection used and the experimental results obtained, while in the last section we summarize our work and discuss some future work directions.

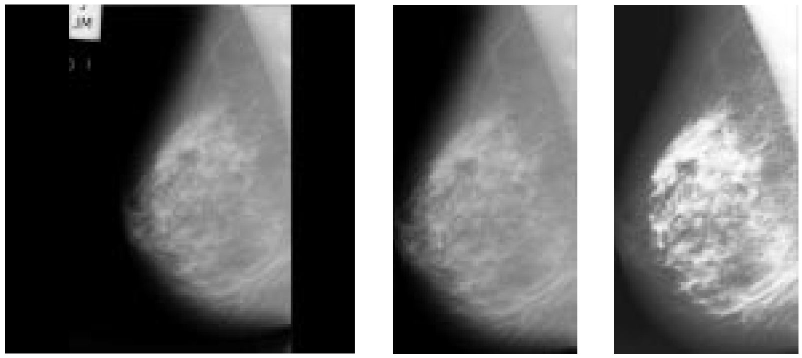
2 Data Cleaning and Feature Extraction

This section summarizes the techniques used to enhance the mammograms as well as the features that were extracted from images. The result of this phase is a transactional database to be mined in the next step of our system. Indeed, we model the images with a set of transactions, each transaction representing one image with the visual features extracted as well as other given characteristics along with the class label.

2.1 Pre-processing phase

Since real-life data is often incomplete, noisy and inconsistent, pre-processing becomes a necessity [12]. Two pre-processing techniques, namely Data Cleaning and Data Transformation, were applied to the image collection. Data Cleaning is the process of cleaning the data by removing noise, outliers etc. that could mislead the actual mining process. In our case, we had images that were very large (typical size was 1024 x 1024) and almost 50% of the whole image comprised the background with a lot of noise. In addition, these images were scanned at different illumination conditions, and therefore some images appeared too bright and some were too dark. The first step toward noise removal was pruning the images with the help of the crop operation in Image Processing. Cropping cuts off the unwanted portions of the image. Thus, we eliminated almost all the background information and most of the noise. An example of cropping that eliminates the artefacts and the black background is given in Figure 1 (a-b).

Since the resulting images had different sizes, the x and the y coordinates were normalized to a value between 0 and 255. The cropping operation was done automatically by sweeping horizontally through the image. The next step towards pre-processing the images was using image enhancement techniques. Image enhancement helps in qualitative improvement of the image with respect to a specific application [10]. Enhancement can be done either in the spatial domain or in the frequency domain. Here we work with the spatial domain and directly deal with the image plane itself. In order to diminish the effect of over-brightness or over-darkness in images, and at the same time accentuate the image features, we applied the Histogram Equalization method, which is a widely used technique. The noise removal step was necessary before this enhancement



(a) (b) (c)

Fig. 1. Pre-processing phase on an example image: (a) original image; (b) crop operation; (c) histogram equalization

because, otherwise, it would also result in enhancement of noise. Histogram Equalization increases the contrast range in an image by increasing the dynamic range of grey levels [10]. Figure 1 (c) shows an example of histogram equalization after cropping.

2.2 Feature Extraction

The feature extraction phase is needed in order to create the transactional database to be mined. The features that were extracted were organized in a database, which is the input for the mining phase of the classification system. The extracted features are four statistical parameters: mean, variance, skewness and kurtosis; the mean over the histogram and the peak of the histogram.

The general formula for the statistical parameters computed is the following:

$$M_n = \frac{\sum (x - \overline{x})^n}{N} \ . \tag{1}$$

where N is the number of data points and n is the order of the moment. The skewness can be defined as:

$$Sk = \frac{1}{N} * \left(\frac{(x - \overline{x})}{\sigma} \right)^3 \ . \tag{2}$$

and the kurtosis as:

$$kurt = \frac{1}{N} * \left(\frac{(x - \overline{x})}{\sigma} \right)^4 - 3 \ . \tag{3}$$

where σ is the standard deviation.

2.3 Transactional Database Organization

All the extracted features presented above have been computed over smaller windows of the original image. The original image was split in four parts, as shown in Figure 2, for a better localization of the region of interest. In addition, the features extracted were discretized over intervals before organizing the transactional data set.

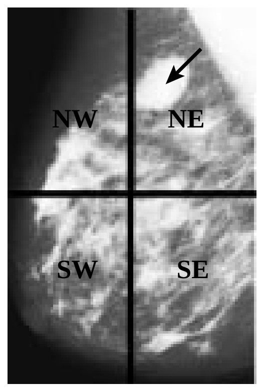


Fig. 2. Mammography division

There are two database organizations that we propose in this paper. The first one is done as follows. The features of all quadrants were kept regardless of whether they characterized normal or cancerous tissue. In addition some other descriptors from the original database were attached, such as breast position, type of tissue, etc.

In the second organization, when all the features were extracted the transactional database to be mined was built in the following way. For the normal images, all the features extracted were attached to the corresponding transaction, while for those characterizing an abnormal mammogram only the features extracted from abnormal quadrants were attached. (e.g. for the mammogram presented in Figure 2 only the features extracted for the NE quadrant (the arrow in the figure points to the tumor) were attached; if the mammogram were a normal one, the features extracted for all the splits would have been attached). In the second organization, in addition to selecting quadrants with tumors from abnormal mammograms, we also dropped those additional features from the database because some of them may not be available in other datasets, while others (breast position) proved to mislead the classification process.

3 Association Rule based Classification

3.1 Association Rules

Association rule mining has been extensively investigated in the data mining literature. Many efficient algorithms have been proposed, the most popular being apriori [1] and FP-Tree growth [11]. Association rule mining typically aims at discovering associations between items in a transactional database. Given a set of transactions $D = \{T_1, .., T_n\}$ and a set of items $I = \{i_1, .., i_m\}$ such that any transaction T in D is a set of items in I, an association rule is an implication $A \rightarrow B$ where the antecedent A and the consequent B are subsets of a transaction T in D, and A and B have no common items. For the association rule to be strong, the conditional probability of B given A has to be higher than a threshold called minimum confidence. Association rules mining is normally a two-step process, where in the first step frequent item-sets are discovered (i.e. item-sets whose support is no less than a minimum support) and in the second step association rules are derived from the frequent item-sets.

In our approach, we used the apriori algorithm in order to discover association rules among the features extracted from the mammography database and the category to which each mammogram belongs. We constrained the association rules such that the antecedent of the rules is composed of a conjunction of features from the mammogram while the consequent of the rule is always the category to which the mammogram belongs. In other words, a rule would describe frequent sets of features per category normal, benign and malign based on the apriori association rule discovery algorithm.

We developed two associative classifiers as described in the following sections.

3.2 Association Rule-based Classification with All Categories

This section introduces the rule generation phase of building an associative classifier when the rules are extracted from the entire training set at once. In this approach (Figure 3) all the transactions in the database from a single training collection and the rules generated are *de facto* the classifier.

The following algorithm presents step by step the process of discovering association rules when the training set is mined at once.

Algorithm ARC-AC Find association rules on the training set of the data collection

Input A set of objects $\mathcal{O}_1$ of the form $\mathcal{O}_i : \{cat_1, cat_2, ...cat_m, f_1, f_2, ...f_n\}$ where cat_i is a category attached to the object and f_j are the selected features for the object; A minimum support threshold σ;

Output A set of association rules of the form $f_1 \wedge f_2 \wedge ... \wedge f_n \Rightarrow cat_i$ where cat_i is a category and f_j is a feature;

Method:

(1) $C_0 \leftarrow \{$Candidate categories and their *support*$\}$
(2) $F_0 \leftarrow \{Frequent$ categories and their *support*$\}$

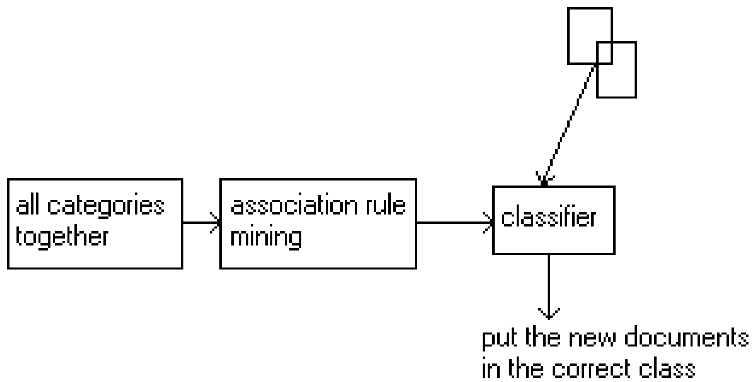


Fig. 3. Classifier for all categories

```
C_1 ← {Candidate 1 itemsets and their support}
F_1 ← {Frequent 1 itemsets and their support}
C_2 ← { candidate pairs (cat, f) such that (cat, f) ∈ 𝒪_1
and cat ∈ F_0 and f ∈ F_1}
foreach object o in 𝒪_1 do {
        foreach c = (cat, f) in C_2 do {
                c.support ← c.support + Count(c, o)
        }
}
F_2 ← {c ∈ C_2 | c.support > σ}
𝒪_2 ← FilterTable(𝒪_1, F_2)
for (i ← 3; F_{i−1} ≠ ∅; i ← i + 1) do {
        C_i ← (F_{i−1} ⋈ F_2)            /* ∀c ∈ C_i c has only one category */
        C_i ← C_i − {c | (i − 1) item-set of c ∉ F_{i−1}}
        𝒪_i ← FilterTable(𝒪_{i−1}, F_{i−1})
        foreach object o in 𝒪_i do {
                foreach c in C_i do {
                        c.support ← c.support + Count(c, o)
                }
        }
        F_i ← {c ∈ C_i | c.support > σ}
}
Sets ← ⋃_i {c ∈ F_i | i > 1}
R= ∅
foreach itemset I in Sets do {
        R ← R + {f ⇒ c | f ∪ c ∈ I ∧ f is an itemset ∧c ∈ C_0}
}
```

Algorithm ARC-AC generates the strong rules when the entire collection is mined. In steps (1-10) the two-frequent itemsets are generated by joining the frequent categories and frequent 1-itemset. Step (11) retains only those that exceed the minimum support threshold. In (13-23) all the k-frequent itemsets are discovered as explained in the apriori algorithm. The last 4 steps represent the actual association rule generation stage. With both algorithms, ARC-AC and ARC-BC (presented next), the document space is reduced in each iteration by eliminating the transactions that do not contain any of the frequent itemsets. This step is done by $\mathit{FilterTable}(\mathcal{O}_{i-1},\ F_{i-1})$ function.

3.3 Association Rule based Classification by Category

This section introduces the second classification method (ARC-BC=association rule based classification by category) that we propose to be applied to the image data collection. It mines the data set by classes instead of mining the entire data set at once. This algorithm was first proposed for text classification in [16].

The transactional database consists of transactions as follows. If an object O_i is assigned to a set of categories $C = \{c_1, c_2, ...c_m\}$ and after preprocessing phase the set of features $F = \{f_1, f_2, ...f_n\}$ is retained, the following transaction is used to model the object: $O_i : \{c_1, c_2, ...c_m, f_1, f_2, ...f_n\}$ and the association rules are discovered from these transactions.

In this approach (Figure 4), each class is considered as a separate training collection and the association rule mining applied to it. In this case, the transactions that model the training documents are simplified to $O_i : \{C, t_1, t_2, ...t_n\}$ where C is the category considered.

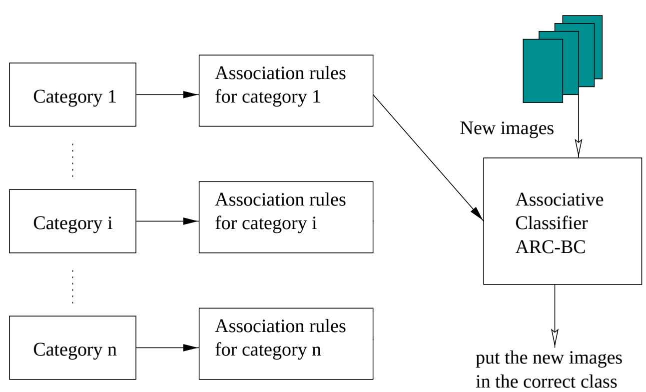


Fig. 4. Classifier per category

In our algorithm we use a constraint so that only the rules that could be used further for classification are generated. In other words, given the transaction model described above, we are interested in rules of the form O $\Rightarrow c_i$ where $O \subseteq O$ and $c_i \subseteq C$. To discover these interesting rules efficiently we push the rule shape constraint in the candidate generation phase of the apriori algorithm in order to retain only the suitable candidate itemsets. Moreover, at the phase for rule generation from all the frequent k-itemsets, we use the rule shape constraint again to prune those rules that are of no use in our classification.

Algorithm ARC-BC Find association rules on the training set of the transactional database when the collection is divided in subsets by category
Input A set of objects (O) of the form $O_i : \{c_i, f_1, f_2, ...f_n\}$ where c_i is the category attached to the object and f_j are the selected features for the object; A minimum support threshold σ; A minimum confidence threshold;
Output A set of association rules of the form $f_1 \wedge f_2 \wedge ... \wedge f_n \Rightarrow c_i$ where c_i is the category and f_j is a feature;
Method:

(1) $C_1 \leftarrow \{$Candidate 1 term-sets and their *support*$\}$
(2) $F_1 \leftarrow \{$Frequent 1 term-sets and their *support*$\}$
(3) for $(i \leftarrow 2; F_{i-1} \neq \emptyset; i \leftarrow i+1)$ $do\{$
(4) $\quad C_i \leftarrow (F_{i-1} \bowtie F_{i-1})$
(5) $\quad C_i \leftarrow C_i - \{c \mid (i-1)$ item-set of $c \notin F_{i-1}\}$
(6) $\quad \mathcal{O}_i \leftarrow \mathit{FilterTable}(\mathcal{O}_{i-1},\ F_{i-1})$
(7) $\quad$ foreach object o in $\mathcal{O}_i$ do $\{$
(8) $\qquad$ foreach c in C_i do $\{$
(9) $\qquad\quad \mathit{c.support} \leftarrow \mathit{c.support} + \mathit{Count}(c,\ o)$
(10) $\qquad \}$
(11) $\quad \}$
(12) $\quad F_i \leftarrow \{c \in C_i \mid \mathit{c.support} > \sigma\}$
(13) $\}$
(14) Sets $\leftarrow \bigcup_i \{c \in F_i \mid i > 1\}$
(15) R$= \emptyset$
(16) foreach itemset I in Sets do $\{$
(17) $\quad R \leftarrow R + \{I \Rightarrow Cat\}$
(18) $\}$

Fig. 5. ARC-BC algorithm

In ARC-BC algorithm step (2) generates the frequent 1-itemset. In steps (3-13) all the k-frequent itemsets are generated and merged with the category in C_1. Steps (16-18) generate the association rules.

4 Building the Classifier

This section describes how the classification system is built and how a new image can be classified using this system. First, there are presented a number of pruning techniques that were considered during our experiments and second, the process of classifying a new image is described.

4.1 Pruning Techniques

The number of rules that can be generated in the association rule mining phase could be very large. There are two issues that must be addressed in this case. The first is that a huge number of rules could contain noisy information which would mislead the classification process. The second is that a huge set of rules would extend the classification time. This could be an important problem in applications where fast responses are required. In addition, in a medical application, it is reasonable to present a small number of rules to medical staff for further study and manual tuning. When the set of rules is too large, it becomes unrealistic to manually sift through it for editing.

The pruning methods that we employ in this project are the following: eliminate the specific rules and keep only those that are general and with high confidence, and prune some rules that could introduce errors at the classification stage. The following definitions introduce the notions used in this subsection.

Definition 1 Given two rules $T_1 \Rightarrow C$ and $T_2 \Rightarrow C$ we say that the first rule is a general rule if $T_1 \subseteq T_2$.

The first step of this process is to order the set of rules. This is done according to the following ordering definition.

Definition 2 Given two rules R_1 and R_2, R_1 is higher ranked than R_2 if:

(1) R_1 has higher confidence than R_2

(2)if the confidences are equal supp(R_1) must exceed supp(R_2)

(3) both confidences and support are equal but R_1 has less attributes in left hand side than R_2

With the set of association rules sorted, the goal is to select a subset that will build an efficient and effective classifier. In our approach we attempt to select a high quality subset of rules by selecting those rules that are general and have high confidence. The algorithm for building this set of rules is described below.

Algorithm Pruning the low ranked specific association rules

Input The set of association rules that were found in the association rule mining phase (S)

Output A set of rules used in the classification process

Method:

(1) sort the rules according to **Definition1**
(2) foreach rule in the set S do {
(2.1) find all those rules that are more specific
(2.2) prune those that have lower confidence
(3) }

The next pruning method employed is to eliminate conflicting rules, rules that for the same characteristics would point to different categories. For example, given two rules $T_1 \Rightarrow C_1$ and $T_1 \Rightarrow C_2$ we say that these are conflicting since they could introduce errors. Since we are interested in a single-class classification, all these duplicates or conflicting rules are eliminated.

4.2 Classifying a new image

The set of rules that were selected after the pruning phase represent the actual classifier. This categorizer is used to predict to which classes new objects are attached. Given a new image, the classification process searches in this set of rules for finding the class that is the closest to be attached with the object presented for categorization. This subsection discusses the approach for labeling new objects based on the set of association rules that forms the classifier.

A solution for classifying new objects is to attach to the new image the class that has the most rules matching this new image or the class associated with the first rule that applies to the new object.

Given an object to classify, the features are extracted and a transaction is create as discussed in Section 2. The features in the object would yield a list of applicable rules in the limit given by the confidence threshold. If the applicable rules are grouped by category in their consequent part and the groups are ordered by the sum of rules' confidences, the ordered groups would indicate the most significant category that should be attached to the object to be classified.

The next algorithm describes the classification of a new image.

Algorithm Classification of a new image (I)

Input A new image to be classified; The associative classifier (ARC); The confidence threshold conf.t;

Output Category attached to the new image

Method:

```
Foreach rule R in ARC (the sorted set of rules) do {
     if R matches I then R.count++ and keep R;
     if R.count==1 then first.conf=R.conf;
     else if (R.conf>first.conf-conf.t) R.count++ and keep R;
     else exit;
}
Let S be the set of rules that match I
Divide S in subsets by category: S_1, S_2...S_n
Foreach subset S_1, S_2...S_n do {
     Sum the confidences of rules in S_k
     put the new document in the class that has the highest confidence sum
}
```

5 Experimental Results

This section introduces the data collection that we used and the experimental results obtained using the two classification methods proposed.

5.1 Mammography Collection

The data collection used in our experiments was taken from the Mammographic Image Analysis Society (MIAS) [14]. Its corpus consists of 322 images, which belong to three categories: normal, benign and malign. There are 208 normal images, 63 benign and 51 malign, which are considered abnormal. In addition, the abnormal cases 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. The existing data in the collection consists of the location of the abnormality (like the centre 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 medio-lateral oblique view. We selected this dataset because it is freely available, and to be able to compare our method with other published work since it is a commonly used database for mammography categorization. We divided the dataset in ten splits to perform the experiments. For each split we selected about 90% of the dataset for training and the rest for testing. That is 288 images in the training set and 34 images in the testing set.

5.2 Experimental Results - Organization 1

In the training phase, the ARC-AC algorithm was applied on the training data and the association rules were extracted. The support was set to 10% and the confidence to 0%. The reason for choosing the 0% percent for the confidence is motivated by the fact that the database has more normal cases (about 70%). The 0% confidence threshold allows us to use the confidence of the rule in the tuning phase of the classifier. In the classification phase, the low and high thresholds of confidence are set such as the maximum recognition rate is reached. The success rate for association rule classifier was 69.11% on average. The results for the ten splits of the database are presented in Table 1. One noticeable advantage of the association rule-based classifier is the time required for training, which is very low compared to other methods such as neural networks.

Given this data organization some experiments were performed with ARC-BC algorithm as well, but the results obtained were unsatisfactory.

5.3 Experimental Results - Organization 2

We have tested our classification approach with ten different splits of the dataset. For Table 2 that is presented below, the association rules are discovered setting a starting minimum support at 25% and the minimum confidence at 50%. The computation of the actual support with which the database is mined is computed in an adaptive way. Starting with the given minimum support the dataset is mined, then a set of association rules is found. These rules are ordered and used as a classifier to test the classifier on the training set. When the accuracy on the training set is higher than a given accuracy threshold, the mining

Table 1. Success ratios for the 10 splits with the association rule based classifier with all categories (ARC-AC)

Database split	Success ration (percentage)
1	67.647
2	79.412
3	67.647
4	61.765
5	64.706
6	64.706
7	64.706
8	64.706
9	67.647
10	88.235
Average	69.11

process is stopped, otherwise the support is decreased ($\sigma = \sigma - 1$) and the process is continued. As a result, different classes are mined at different supports. The parameters in the tests with the results below are: minimum support 25%, minimum confidence 50% and the accuracy threshold is 95%.

Table 2. Classification accuracy over the 10 splits using ARC-BC

Split	1st rule		ordered		cut rules		remove specific	
	#rules	accuracy	#rules	accuracy	#rules	accuracy	#rules	accuracy
1	22	76.67	1121	80.00	856	76.67	51	60.00
2	18	86.67	974	93.33	755	90.00	48	86.67
3	22	83.33	823	86.67	656	86.67	50	76.67
4	22	63.33	1101	76.67	842	66.67	51	53.33
5	33	56.67	1893	70.00	1235	70.00	63	50.00
6	16	66.67	1180	76.67	958	73.33	51	63.33
7	30	66.67	1372	83.33	1055	73.33	58	53.33
8	26	66.67	1386	76.67	1089	80.00	57	46.67
9	20	66.67	1353	76.67	1130	76.67	52	60.00
10	18	76.67	895	83.33	702	80.00	51	76.67
avg(%)	22.7	**71.02**	1209.8	**80.33**	927.8	**77.33**	53.2	**62.67**

Classification in the first two columns of Table 2 is done by assigning the image to the category attached to the first rule (the one with the highest confidence) that applies to the test image (see Table 2 columns under '1st rule'). However, pruning techniques are employed before so that a high quality set of rules is selected. The pruning technique used in this case is a modified version of the database coverage (i.e. selecting a set of rules that classifies most transactions presented in the training set). Given a set of rules, the main idea is to find the best rules that would make a good distinction between the classes. The given set of rules is ordered. Take one rule at a time and classify the training set for each class. If the consequent of the rule indicates class c_i keep that rule, only if it correctly classifies some objects in c_i training set and doesn't classify any in the other classes. The transactions that were classified are removed from the training set.

The next columns in Table 2 are results of classification that uses the most powerful class in the set of rules. The difference is as follows: in the first two columns the set of rules that form the classifier is the set of rules extracted at the mining stage but ordered according to the confidence and support of the rules (support was normalized so that the ordering is possible even if the association rules are found by category)(see Table 2 columns under 'ordered'); in the next two columns after the rules were ordered the conflicting rules (see Section 4.1) were removed (see Table 2 columns under 'cut rules'); in the last two columns (see Table 2 columns under 'remove specific') from the ordered set of rules the specific ones were removed if they had lower confidence (see Section 4.1).

We also present precision/recall graphs in Figure 6 to show that both false positive and false negative are very small for normal cases, which means that for abnormal images we have a very small number of false negative which is very desirable in medical image classification.

The formulae for precision and recall are given below:

$$R = \frac{TP}{TP + FN} \ . \tag{4}$$

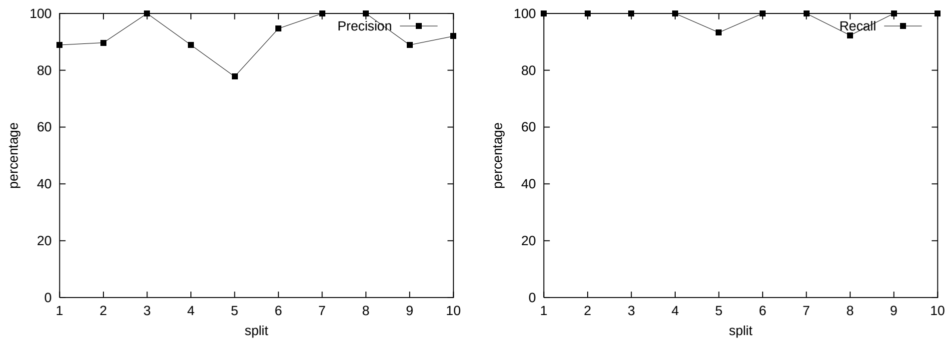
$$P = \frac{TP}{TP + FP} \ . \tag{5}$$

The terms used to express precision and recall are given in the contingency table Table 3, where TP stands for true positives, FP for false positives, FN for false negatives and TN for true negatives.

From the graphs presented in Figure 6 one can observe that for both precision and recall for normal cases the values are very high. In addition, we can notice from equations 4 and 5 that the values for FP and FN tend to zero when precision and recall tend to 100%. Thus, the false positives and in particular false negatives are almost null with our approach.

Table 3. Contingency table for category *cat*

Category *cat*		human assignments: Yes	human assignments: No
classifier assignments	Yes	TP	FP
classifier assignments	No	FN	TN



(a) (b)

Fig. 6. (a)Precision over the ten splits ; (b) Recall over the ten splits

In Table 4 the classification is done using the association rules obtained when mining the entire dataset at once with the second organization. In the first two columns the set of rules that form the classifier is the set of rules extracted at the mining stage but ordered according to the confidence and support of the rules (see Table 4 columns under 'ordered'); in the next two columns after the rules were ordered the conflicting rules (see Section 4.1) were removed (see Table 4 columns under 'cut rules').

Table 4. Classification accuracy over the 10 splits using ARC-AC[2]

Split	ordered		cut rules	
	#rules	accuracy	#rules	accuracy
1	6967	53.33	6090	53.33
2	5633	86.67	4772	86.67
3	5223	76.67	4379	76.67
4	6882	53.33	5938	53.33
5	7783	50.00	6878	50.00
6	7779	60.00	6889	60.00
7	7120	46.67	6209	46.67
8	7241	43.33	6364	43.33
9	7870	53.33	6969	53.33
10	5806	76.67	4980	76.67
avg(%)	6830.4	**60.00**	5946.8	**60.00**

As observed from the two tables presented above, the accuracy reached when ARC-BC is used is higher than the one obtained when the training set was mined at once with ARC-AC. However, the accuracy reached in [2] with ARC-AC was actually higher than in this case (69.11%). These results prove the importance of choosing the right data cleaning technique and data organization in reaching an effective and efficient data mining system.

Not only in accuracy does ARC-BC outperform ARC-AC, but in time measurements as well (41.315 seconds versus 199.325 seconds for training and testing for all ten splits). All tests were performed on an AMD Athlon 1.8 GHz.

6 Conclusions

In this paper we presented two classification methods applied to medical image classification. Both classification systems are based on association rule mining. In addition, we demonstrated how important the cleaning phase is in building a classification system. The evaluation of the system was carried out on MIAS [14] dataset and the experimental results show that the accuracy of the system reaches 80.33% accuracy and the false negatives and false positives tend towards zero in more than half the splits. Although the results seem promising when an associative classifier is used, there are some future research directions to be studied. A collaboration with medical stuff would be very interesting in order to evaluate the performance of our system. In addition, the extraction of different features or a different database organization could lead to improved results.

References

1. R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In *Proc. 1993 ACM-SIGMOD Int. Conf. Management of Data*, pages 207–216, Washington, D.C., May 1993.
2. Maria-Luiza Antonie, Osmar R. Zaïane, and Alexandru Coman. Application of data mining techniques for medical image classification. In *In Proc. of Second Intl. Workshop on Multimedia Data Mining (MDM/KDD'2001) in conjunction with Seventh ACM SIGKDD*, pages 94–101, San Francisco, USA, 2001.
3. D. Brazokovic and M. Neskovic. Mammogram screening using multiresolution-based image segmentation. *International Journal of Pattern Recognition and Artificial Intelligence*, 7(6):1437–1460, 1993.
4. C. Chen and G. Lee. Image segmentation using multiresolution wavelet analysis and expectation-maximization (em) algorithm for digital mammography. *International Journal of Imaging Systems and Technology*, 8(5):491–504, 1997.
5. A. Dhawan et al. Radial-basis-function-based classification of mammographic microcalcifications using texture features. In *Proc. of the 17th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 1, pages 535–536, 1995.
6. H. Chan et al. Computerized analysis of mammographic microcalcifications in morphological and feature spaces. *Medical Physics*, 25(10):2007–2019, 1998.
7. H. Li et al. Markov random field for tumor detection in digital mammography. *IEEE Trans. Medical Imaging*, 14(3):565–576, 1995.
8. H. Li et al. Fractal modeling and segmentation for the enhancement of microcalcifications in digital mammograms. *IEEE Trans. Medical Imaging*, 16(6):785–798, 1997.

9. I. Christoyianni et al. Fast detection of masses in computer-aided mammography. *IEEE Signal Processing Magazine*, pages 54–64, 2000.
10. Rafael C. Gonzalez and Richard. E. Woods. *Digital Image Processing*. Addison-Wesley, 1993. second edition.
11. J. Han, J. Pei, and Y. Yin. Mining frequent patterns without candidate generation. In *ACM-SIGMOD*, Dallas, 2000.
12. Jiawei Han and Micheline Kamber. *Data Mining, Concepts and Techniques*. Morgan Kaufmann, 2001.
13. S. Lai, X. Li, and W. Bischof. On techniques for detecting circuscribed masses in mammograms. *IEEE Trans. Medical Imaging*, pages 377–386, 1989.
14. http://www.wiau.man.ac.uk/services/MIAS/MIASweb.html.
15. T. Wang and N. Karayiannis. Detection of microcalcification in digital mammograms using wavelets. *IEEE Trans. Medical Imaging*, pages 498–509, 1998.
16. Osmar R. Zaïane and Maria-Luiza Antonie. Clasifying text documents by associating terms with text categories. In *In Proc. of the Thirteenth Australasian Database Conference (ADC'02)*, pages 215–222, Melbourne, Australia, 2002.