Proceedings of the Second International Workshop on Multimedia Data Mining (MDM/KDD'2001), in conjunction with ACM SIGKDD conference. San Francisco, USA, August 26, 2001

Application of Data Mining Techniques for Medical Image Classification

Maria-Luiza Antonie
Database Laboratory
Department of Computing Science
University of Alberta
Canada
email: luiza@cs.ualberta.ca

Osmar R. Zaïane
Database Laboratory
Department of Computing Science
University of Alberta
Canada
email: zaiane@cs.ualberta.ca

Alexandru Coman
Database Laboratory
Department of Computing Science
University of Alberta
Canada
email: acoman@cs.ualberta.ca

ABSTRACT
Breast cancer represents the second leading cause of cancer deaths in women today and it is the most common type of cancer in women. This paper presents some experiments for tumour detection in digital mammography. We investigate the use of different data mining techniques, neural networks and association rule mining, for anomaly detection and classification. The results show that the two approaches performed well, obtaining a classification accuracy reaching over 70% percent for both techniques. Moreover, the experiments we conducted demonstrate the use and effectiveness of association rule mining in image categorization.

KEYWORDS
classification, medical imaging, association rule mining, neural networks, image categorization, image mining.

1. Introduction

The high incidence of breast cancer in women, especially in developed countries, has increased significantly in the last years. Though much less common, breast cancer also occurs in men [1][15, 14]. 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, which 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 breast cancer. Due to the high volume of mammograms 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 the medical staff to achieve high efficiency and effectiveness.

The methods proposed in this paper classify the digital mammograms in two categories: normal and abnormal. The normal ones are those characterizing a healthy patient. The abnormal ones include both benign cases, representing mammograms showing a tumour that is not formed by cancerous cells, and malign cases, those mammograms taken from patients with cancerous tumours. 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, in 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. Due to the significance of an automated image categorization to help physicians and radiologists, much research in the field of medical images classification has been done recently [16, 20, 9]. With all this effort, there is still no widely used method to classifying medical images. This is due to the fact that the medical domain requires high accuracy and especially the rate of false negatives to be very low. In addition, another important factor that influences the success of 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. Mammography alone cannot prove that a suspicious area is malignant or benign. To decide that, 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 percentages of breast biopsy cases are proved cancerous. In a false negative detection, an actual tumour remains undetected that could lead to higher costs or even to the cost of a human life. Here is the trade-off that appears in developing a classification system that could directly affect human life. In addition, the tumours existing are of different types. Tumours are of different shapes and some of them have the characteristics of the normal tissue. All these reasons make the decisions that

[1] In the United States, for example, male breast cancer accounts for 1 of every 100 cases of breast cancers [15]

are made on such images even more difficult.

Different methods have been used to classify and/or detect anomalies in medical images, such as wavelets [3, 20], fractal theory [8], statistical methods [6] and most of them used features extracted using image-processing techniques [16]. In addition, some other methods were presented in the literature based on fuzzy set theory [2], Markov models [7] and neural networks [9, 5]. 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.

In this paper, we use a common classification method, namely neural networks, but significantly improve the accuracy rate of the classifier compared to other published results using the same data set. In addition, we investigate the use of association rules, typically used in market basket analysis, in the problem of image categorization and demonstrate with encouraging results that association rule mining is a promising alternative in medical image classification and certainly deserves further attention. To the best of our knowledge, association rules have never been used for image categorization. Some research work was published showing the use of FP-growth algorithm [11] for building classifiers [17]. We have also studied text categorization with association rules [21].

The rest of the paper is organized as follows: Section 2 depicts the general classification process, presents the data collection used for benchmarking and describes the image pre-processing phase. Feature extraction is also presented in Section 2. Classification of images using neural networks is presented in Section 3 and classification of images with association rules is introduced in Section 4. In Section 5, we discuss our experiments and the results. Conclusions are presented in Section 6.

2. Data Collection and Preprocessing

To automatically categorize medical images, we have experimented on real mammograms with two data mining techniques, association rule mining and neural networks. In both cases, the problem consists of building a mammography classification model using attributes extracted from and attached to mammograms, then evaluating the effectiveness of the model using new images. The process of building the classification model (classifier) includes pre-processing and extraction of visual features from already labelled images (i.e. training set).

Figure 1 shows an overview of the categorization process adopted for both systems. The first step is represented by the image acquisition and image enhancement, followed by feature extraction. The last one is the classification part where the technique for supervised learning is different. All these parts of the classification systems are discussed in more detail later.

2.1 Mammography Data Collection

To have access to real medical images for experimentation is a very difficult undertaking due to privacy issues and heavy bureaucratic hurdles. The data collection that was used in our experiments was taken from the Mammographic Image Analysis Society (MIAS) [18]. This same collection has been used in other studies of automatic mammography classification. Its corpus consists of 322 images, which belong to three big 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 in 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 tumour), its radius, breast position (left or right), type of breast tissues (fatty, fatty-glandular and dense) and tumour type if exists (benign or malign). All the mammograms are medio-lateral oblique view.

2.2 Pre-processing Phase

Mammograms are images difficult to interpret, and a pre-processing phase of the images is necessary to improve the quality of the images and make the feature extraction phase more reliable. Pre-processing is always a necessity whenever the data to be mined in noisy, inconsistent or incomplete and pre-processing significantly improves the effectiveness of the data mining techniques [12]. This section introduces the pre-processing techniques applied to the images before the feature extraction phase. In the digitization process, noise could be introduced that needs to be reduced by applying some image processing techniques. In addition, at the time that the mammograms were taken, the conditions of illumination are generally different.

We applied to the images two techniques: a cropping operation and an image enhancement one. The first one was employed in order to cut the black parts of the image as well as the existing artefacts such as written labels etc. For most of the images in our dataset, almost 50% of the whole image comprised of a black background with significant noise. Cropping removed the unwanted parts of the image usually periferal to the area of interest. An example of cropping that eliminates the artefacts and the black background is given in Figure 4.

The cropping to eliminate noise was done first before the image enhansement to avoid enhancing noise and hindering the cleaning phase. The cropping operation was done automatically by sweeping through the image and cutting horizontally and vertically the image those parts that had the mean less than a certain threshold.

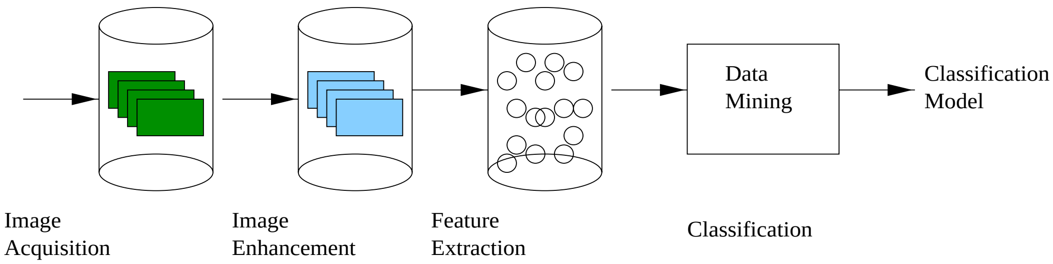

Figure 1. Image categorization process.

Image enhancement helps in qualitative improvement of the image with respect to a specific application [10]. In order to diminish the effect of over brightness or over darkness in the images and accentuate the image features, we applied a widely used technique in image processing to improve visual appearance of images known as Histogram Equalization. Histogram equalization increases the contrast range in an image by increasing the dynamic range of grey levels (or colours) [10]. This improves the distinction of features in the image. The method proceeds by widening the peaks in the image histogram and compressing the valleys. This process equalizes the illumination of the image and accentuates the features to be extracted. That is how the different illumination conditions at the scanning phase are reduced. Figure 4 shows the result of histogram equalization on the cut image.

2.3 Feature Extraction

After cropping and enhancing the images, which represents the data cleaning phase, features relevant to the classification are extracted from the cleaned images. The extracted features are organized in a database in the form of transactions, which in turn constitute the input for both classification algorithms used. The transactions are of the form $\{$ImageID, Class Label, $F_1, F_2, ..., F_n\}$ where $F_1 ... F_n$ are ne n features extracted for a given image. This database is constructed by merging some already existing features in the original database with some new visual content features that we extracted from the medical images using image-processing techniques. The existing features are:

- The type of the tissue (dense, fatty and fatty-glandular);
- The position of the breast: left or right.

The type of tissue is an important feature to be added to the feature database, being well known the fact that for some types of tissue the recognition is more difficult than for others. Training the classification systems with these features incorporated could increase the accuracy rate. The extracted features are four statistical parameters:

1. mean;
2. variance;
3. skewness and
4. kurtosis.

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

$$M_n = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^n}{N}$$

where N is the number of data points, and n is the order of the moment.

The skewness can be defined as:

$$Sk = \frac{1}{N}\left(\frac{(x-\bar{x})}{\sigma}\right)^3$$

and the kurtosis as :

$$Kurt = \frac{1}{N}\left(\frac{(x-\bar{x})}{\sigma}\right)^4 - 3$$

All these extracted features are computed over smaller windows of the original image. The original image is first split in four parts as is shown in Figure 2. For a more accurate extraction of the features and for a further investigation of the localisation we split each of these four regions in other four parts. The statistical parameters were computed for each of the sixteen sub-parts of the original image.

3. Neural Networks

3.1 Theoretical Background

Artificial neural network models have been studied for many years in the hope of achieving human-like performance in several fields such as speech and image understanding. The networks are composed of many non-linear computational elements operating in parallel and arranged in patterns reminiscent of biological neural networks. Computational elements or nodes are connected in

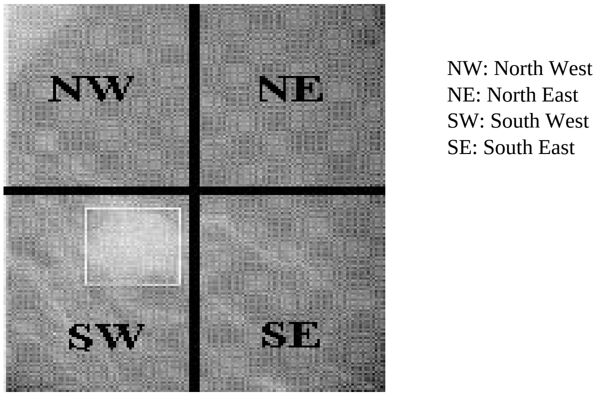

Figure 2. The four regions of the first division, and then, for each of the areas is further divided in four.

several layers (input, hidden and output) via weights that are typically adapted during the training phase to achieve high performance. Instead of performing a set of instructions sequentially as in a Von Neumann computer, neural network models explore simultaneously many hypotheses using parallel networks composed of many computational elements connected by links with variable weights.

The back-propagation algorithm is an extension of the least mean square (LMS) algorithm that can be used to train multi-layer networks. Both LMS and back-propagation are approximate steepest descent algorithms that minimize squared error. The only difference between them is in the way in which the gradient is calculated. The back-propagation algorithm uses the chain rule in order to compute the derivatives of the squared error with respect to the weights and biases in the hidden layers. It is called back-propagation because the derivatives are computed first at the last layer of the network, and then propagated backward through the network, using the chain rule, to compute the derivatives in the hidden layers. For a multi-layer network, the output of one layer becomes the input of the following layer. A typical 2-layer neural network is depicted in Figure 3. It schematizes the neural network we used with one node in the output layer since we aimed at two class labels only.

In the following sections, we shall describe the details of our architecture.

3.2 The architecture of the neural network based system

The architecture of the neural network consists of three layers: an input layer, a hidden one and an output layer. The number of nodes in the input layer is equal to the number of elements existing in one transaction in the database. In our case, the input layer had 69 nodes. For the hidden layer, we chose 10 nodes, while the output layer was consisting of

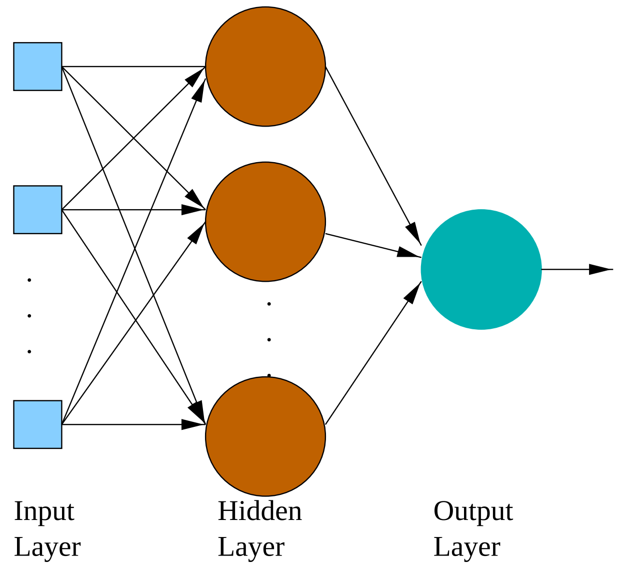

Figure 3. A 2-layer neural network.

one node. The node of the output layer is the one that gives the classification for the image. It classifies it as normal or abnormal.

In the training phase, the internal weights of the neural network are adjusted according to the transactions used in the learning process. For each training transaction the neural network receives in addition the expected output. This allows the modification of the weights. In the next step, the trained neural network is used to classify new images.

4. Association Rule Mining

4.1 Theoretical Background

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 acceptable, 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, wherein 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 to be discovered 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 and abnormal (benign and malign) based on the apriori association rule discovery algorithm.

After all the features are merged and put in the transactional database, the next step is applying the apriori algorithm for finding the association rules in the database constrained as described above with the antecedent being the features and the consequent being the category. Once the association rules aree found, they are used to construct a classification system that categorizes the mammograms as normal, malign or benign. The most delicate part of the classification with association rule mining is the construction of the classifier itself. Although we have the knowledge extracted from the database by finding the existing association rules, the main question is how to build a powerful classifier from these associations. The association rules that have been generated from the database in such a manner that they have as consequent a category from the classification classes. The association rules could imply either normal or abnormal. When a new image has to be classified, the categorization system returns the association rules that applies to that image. The first intuition in building the classification system is to categorize the image in the class that has the most rules that apply. This classification would work when the number of rules extracted for each class is balanced. In other cases, a further tuning of the classification system is required. The tuning of the classifier is mainly represented by finding some optimal intervals of the confidence such as both the overall recognition rate and the recognition rate of abnormal cases are at its maximum value. In dealing with medical images it is very important that the false negative rate be as low as possible. It is better to misclassify a normal image than an abnormal one. That is why in our tuning phase we take into consideration the recognition rate of abnormal images. It is not only important to recognize some images, but to be able to recognize those that are abnormal.

By applying the apriori algorithm with additional constraints on the form of the rules to be discovered we generate a relatively small set of association rules associating sets of features with class labels. These association rules constitute our classification model. The discovery of association rules in the mammogram feature database represents the training phase of our classifier. Generating the constrained association rules is very fast by comparison with training a neural network. To classify a new mammogram, it suffises to extract the features from the image as was done for the training set, and applying the association rules on the extracted features to identify the class the new mammogram falls into.

5. Experimental Results

In our experiments, we considered the 322 images from the database for both classification systems. From these set of images we considered 90 percent for training the systems and 10 percent for testing them. We considered ten splits of the data collection and computed the results for all of them in order to obtain a more accurate result of the systems' potential.

5.1 Neural Networks

The results obtained using the neural network as classifier are presented in Table 1.

On average, the classifier performed extremely well compared to other methods presented in the literature. However, the classification success ratio was not consistent among the different splits and ranged from 65.6% for split 7 to 93.7% for split 10. This inconsistency makes the method nonviable in real life applications. Nevertheless, even the lowest success rate of 65.6% can be a significant helper for a physician as an initial categorization.

Database split	Success ration (percentage)
1	96.870
2	90.620
3	90.620
4	78.125
5	81.250
6	84.375
7	65.625
8	75.000
9	56.250
10	93.750
	Average: 81.248

Table 1. Success ratios for the 10 splits with the neural network based classifier.

5.2 Association Rule Mining

As in any learning process for building a classifier, the classification performed with association rule mining comprised two steps. The first one is represented by the training of the system, while the second one deals with the classification of the new images.

In the training phase, the apriori algorithm was applied on the training data and the association rules were

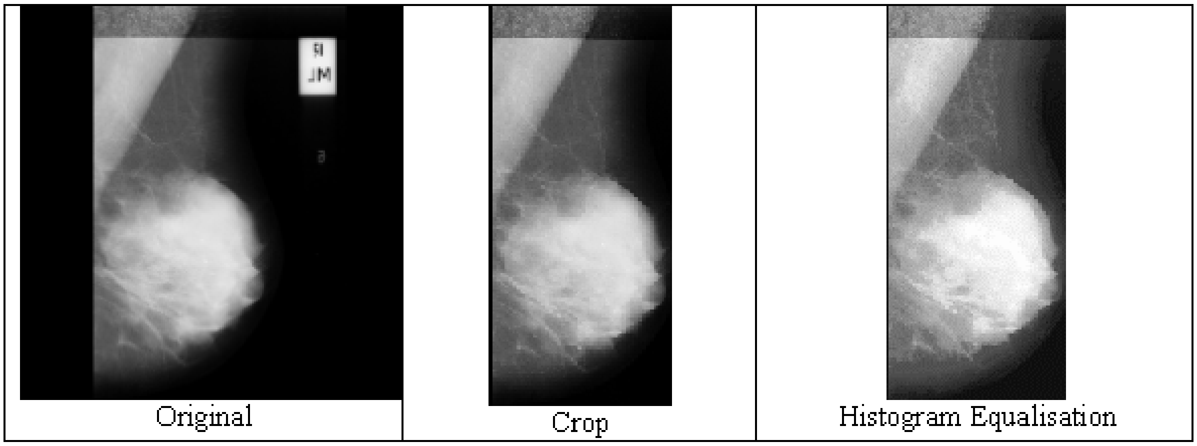

Figure 4. Pre-processing phase on an example image.

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 2. 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.

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

Table 2. Success ratios for the 10 splits with the association rule based classifier.

The recognition rate obtained using association rule mining is close to some other results reported in the literature. Another interested fact to be noticed is that the classifier proves to perform well on all the splits of the database, being more compact and consistent than the neural network classifier.

We noticed that the association rule classifier was sensitive to the unbalanced data collection that contained about 70 percent normal cases and only 30 percent abnormal, this being further divided into benign and malign. This is why we decided to build another classifier using an equilibrated distribution of normal and abnormal cases. For comparison reasons, we used a split that was also chosen in [9]. The same split is not the only reason for choosing [9] as comparison. In addition, the feature extraction phase is similar and a radial basis function network represents the classifier. We considered the 22 mammograms containing circumscribed lesions existing in the database. From these 22 mammograms, there are 18 benign and 4 malign. The abnormal mammograms are further split according to tissue type in fatty (11 cases), fatty-glandular (8 cases) and dense (3 cases). For the training procedure, we have selected 22 abnormal images and 22 normal images selected at random. For the evaluation of the results, we have used all the abnormal mammograms from MIAS database containing circumscribed masses and another 22 normal mammograms randomly selected. For this split the success rate was better (78.69%) than the previous splits. A noticeable fact is, that due to the imbalance between benign and malign images, the number of rules generated for the malign one was extremely reduced thus all the malign images being misclassified. Three out of four malign images were classified as abnormal (benign) which means that the classification in just normal and abnormal categories was actually higher (84.09%) which is a significant improvement over the results presented in [9] (75.2%). The classification results in normal and abnormal categories are presented in Figure 5.

As compared to the results presented in [9] we obtained a lower recognition rate for the fatty abnormal mammograms, but higher for all the normal cases.

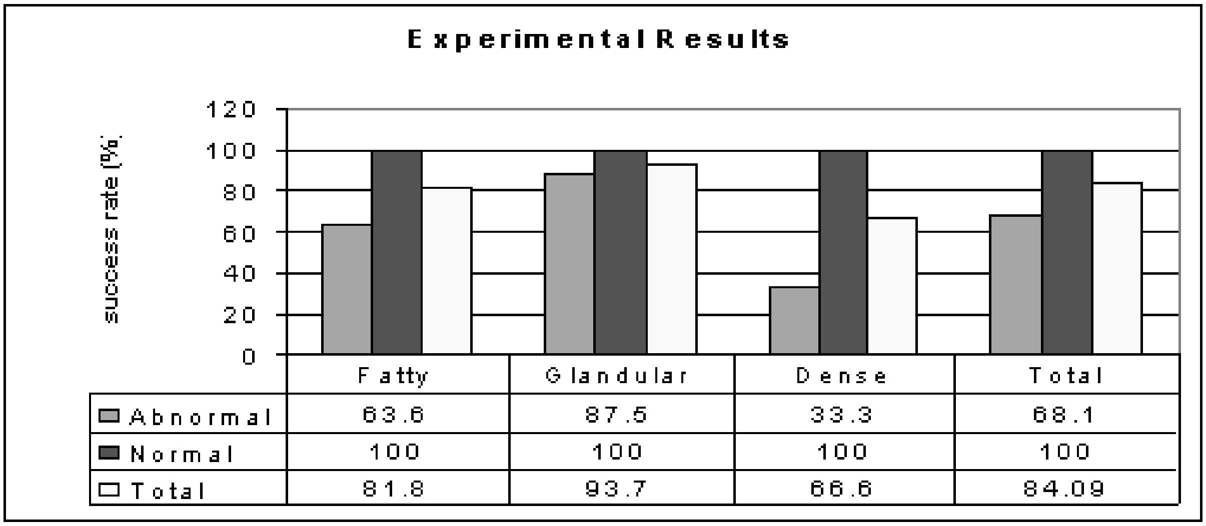

Figure 5. Success rates of association rule mining classifier.

6. Conclusions and Future Work

Mammography is one of the best methods in breast cancer detection, but in some cases, radiologists cannot detect tumours despite their experience. Such computer-aided methods like those presented in this paper could assist medical staff and improve the accuracy of detection.

In this paper, we presented two methods for tumour classification in mammograms. One system exploited the use of neural networks using back-propagation and the second one was built employing association rule mining with constraint form. The first method proved to be less sensitive to the database imbalance at a cost of high training times. The second one, with a much more rapid training phase, obtained better results than reported in literature on a well balance dataset. Both methods performed well which proves that association rules mining employed in classification process is worth further investigation.

It is well known that data mining techniques are more suitable to larger databases than the one used for these preliminary tests. We intend to use a larger mammographic database and to extract more features from the images. In particular, a classification model based on association rules becomes more accurate with a larger dataset than in the order of 300 images. In addition, more features from the database, in particular non-visual features attached to the images such as age, with/without children etc., could be interesting and relevant as additional attributes for classification. We intend to study the influence on the performance of those added features. In the case of the association rule mining approach, image split in more windows than we used could improve the detection by better localization of the cancerous tumour, thus more specific rules being extracted. For the neural network, once we use a larger database we intend to use more sophisticated neural networks in order to reduce the training times and improve accuracy. It has also been observed that the techniques employed for pre-processing the images can significantly improve or worsen the accuracy of the classifier. This was the reason our neural network performed better on the same dataset than other published research also using neural networks. We have also investigated techniques for segmenting the mammograms to determine regions of interest (not reported in this paper). Such segmentations can isolate specific regions that may be of interest to physicians [19]. We have used single link region growing algorithm [13, 4] to segment the image in regions of interest and used the features of eash region as attributes of the image. Unfortunately, this technique while reaching very encouraging accuracy with the association rule mining approach (better than the results reported in this paper) didn't perform as we hoped with the neural network approach. This, yet again, emphasized the importance of image pre-processing and the techniques used for visual feature extraction in the process of multimedia data mining. The pre-processing of mammography and the extraction of features should be dictated by rules that make sense medically. This is one of our future goals to validate the feature extraction by radiologists.

7. Acknowledgement

The authors would like to thank Veena Sridhar for her collaboration in the initial experiments of image classification using neural networks and for her contributions in the image pre-processing by segmentation (not presented in this final paper).

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] 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.

[3] 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.

[4] M. L. Comer, S. Liu, and E. J. Delp. Statistical segmentation of mammograms. In *Proc. of the 3nd International Workshop on Digital Mammography*, pages 475–478, Chicago, June 9-12 1996.

[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. Marcov 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, Jan 2000.

[10] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing, 2nd edition*. Addison-Wesley, 1993.

[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] R. Haralick and L. Shapiro. *Computer and Robot Vision*. Addison-Wesley, Reading, Mass., 1992.

[14] Breast Cancer in Men. A complete patient's guide. http://www.breastdoctor.com/breast/men/cancer.htm.

[15] Breast Cancer in Men. Male breast cancer information center. http://interact.withus.com/interact/mbc/.

[16] S. Lai, X. Li, and W. Bischof. On techniques for detecting circumscribed masses in mammograms. *IEEE Trans. Medical Imaging*, 8(4):377–386, 1989.

[17] George Wenmin Li. Classification based on multiple association rules. Master's thesis, Computing Science, Simon Fraser University, 2001.

[18] http://www.wiau.man.ac.uk/services/MIAS/MIASweb.html.

[19] S. Singh and R. Al-Mansoori. Identification of regions of interest in digital mammograms. *Journal of Intelligent Systems*, 10(2):183–217, 2000.

[20] T. Wang and N. Karayiannis. Detection of microcalcification in digital mammograms using wavelets. *IEEE Trans. Medical Imaging*, 17(4):498–509, 1998.

[21] Osmar R. Zaïane and Maria-Luiza Antonie. Classifying text documents by associating terms with text categories. In *submitted to the Tenth International Conference on Information and Knowledge Management (CIKM 2001)*, Atlanta, GA, USA, November 2001.