International Journal of Computer and Communication Technology

Volume 3 | Issue 1 Article 9

January 2012

Efficient Image Mining Technique for Classification of Mammograms to Detect Breast Cancer

Aswini Kumar Mohanty
SOA University, Khandagiri Bhubaneswar, asw_moh@yahoo.com

Saroj Kumar Lenka
Deptt. Of Computer Science, Modi Univesity,Rajasthan,India., lenka.sarojkumar@gmail.com

Recommended Citation

Mohanty, Aswini Kumar and Lenka, Saroj Kumar (2012) "Efficient Image Mining Technique for Classification of Mammograms to Detect Breast Cancer," *International Journal of Computer and Communication Technology*: Vol. 3 : Iss. 1 , Article 9.
DOI: 10.47893/IJCCT.2012.1113
Available at: https://www.interscience.in/ijcct/vol3/iss1/9

Efficient Image Mining Technique for Classification of Mammograms to Detect Breast Cancer

Aswini Kumar Mohanty,
Deptt. Of Computer Science,
Gandhi Engineering College,
Bhubaneswar – 752 054, Orissa, India
Email : asw_moh@yahoo.com

Saroj Kumar Lenka,
Deptt. Of Computer Science,
Modi Univesity , Lakshmangarh-332311,
Rajasthan, India
Email : lenka.sarojkumar@gmail.com

***Abstract*--The image mining technique deals with the extraction of implicit knowledge and image with data relationship or other patterns not explicitly stored in the images. It is an extension of data mining to image domain. The main objective of this paper is to apply image mining in the domain such as breast mammograms to classify and detect the cancerous tissue. Mammogram image can be classified into normal, benign and malignant class. Total of 24 features including histogram intensity features and GLCM features are extracted from mammogram images. A hybrid approach of feature selection is proposed which approximately reduces 75% of the features and new decision tree is used for classification. Experiments have been taken for a data set of 300 images taken from MIAS of different types with the aim of improving the accuracy by generating minimum no. of rules to cover more patterns.**

Key word— *Mammogram, GLCM feature, Histogram Intensity, Genetic Algorithm, Branch and Bound technique, Decision tree Classification.*

1. INTRODUCTION

Breast Cancer is one of the most common cancers, leading to cause of death among women, especially in developed countries. There is no primary prevention since cause is still not understood. So, early detection of the stage of cancer allows treatment which could lead to high survival rate. Mammography is currently the most effective imaging modality for breast cancer screening. However, 10-30% of breast cancers are missed at mammography [1]. Mining information and knowledge from large database has been recognized by many researchers as a key research topic in database system and machine learning Researches that use data mining approach in image learning can be found in [2-8].

Data mining of medical images is used to collect effective models, relations, rules, abnormalities and patterns from large volume of data. This procedure can accelerate the diagnosis process and decision-making. Different methods of data mining have been used to detect and classify anomalies in mammogram images such as wavelets [18,22], statistical methods and most of them used feature extracted using image processing techniques [5].Some other methods are based on fuzzy theory [1] and neural networks [19],

In this paper we have used classification method called Decision tree classifier for image classification.

Classification process typically involves two phases: training phase and testing phase. In training phase the properties of typical image features are isolated and based on this training class is created .In the subsequent testing phase , these feature space partitions are used to classify the image. We have used supervised decision tree method by extracting low level image features for classification. The merits of this method are effective feature extraction, selection and efficient classification. The rest of the paper is organized as follows. Section 2 presents the preprocessing and section 3 presents the feature extraction phase. Section 4 discusses the proposed method of Feature selection and classification. In section5 the results are discussed and conclusion is presented in section 6.

2. PRE-PROCESSING

The mammogram image for this study is taken from Mammography Image Analysis Society (MIAS), which is an UK research group organization related to the Breast cancer investigation. As mammograms are difficult to interpret, preprocessing is necessary to improve the quality of image and make the feature extraction phase as an easier and reliable one. The calcification cluster/tumor is surrounded by breast tissue that masks the calcifications preventing accurate detection and shown in Figures 2.1 .A pre-processing; usually noise-reducing step is applied to improve image and calcification contrast.

In this work an efficient filter referred to as the low pass filter, was applied to the image that maintained calcifications while suppressing unimportant image features.

Figures 2 shows representative output image of the filter for a image cluster in figure 1. By comparing the two images, we observe background mammography structures are removed while calcifications are preserved. This simplifies the further tumor detection step.

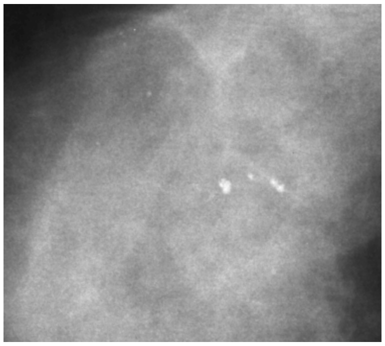

Fig. 1 ROI of a Benign

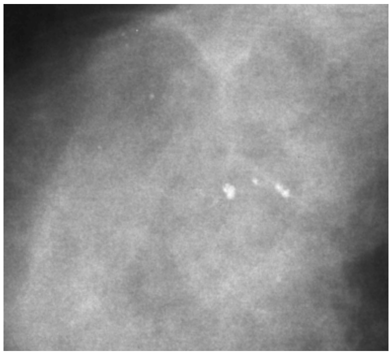

Fig. 2 ROI after Pre-processing Operation

2.1 Histogram Equalization

Histogram equalization is a method in image processing of contrast adjustment using the image's histogram. Through this adjustment, the intensities can be better distributed on the histogram. This allows for areas of lower local contrast to get better contrast. Histogram equalization accomplishes this by efficiently spreading out the most frequent intensity values. The method is useful in images with backgrounds and foregrounds that are both bright or both dark. In particular, the method can lead to better views of bone structure in x-ray images, and to better detail in photographs that are over or under-exposed. In mammogram images Histogram equalization is used to make contrast adjustment so that the image abnormalities will be better visible.

3. FEATURE EXTRACTION

Features, characteristics of the objects of interest, if selected carefully are representative of the maximum relevant information that the image has to offer for a complete characterization a lesion. Feature extraction methodologies analyze objects and images to extract the most prominent features that are representative of the various classes of objects. Features are used as inputs to classifiers that assign them to the class that they represent. In this Work intensity histogram features and Gray Level Co-Occurrence Matrix (GLCM) features are extracted.

3.1 INTENSITY HISTOGRAM FEATURES

Intensity Histogram analysis has been extensively researched in the initial stages of development of this algorithm. Prior studies have yielded the intensity histogram features like mean, variance, entropy etc. These are summarized in Table 3.1(a) Mean values characterize individual calcifications; Standard Deviations (SD) characterize the cluster. Table 3.1(b) summarizes the values for those features.

Table 3.1(a) Intensity histogram features

Feature Number assigned	Feature
1.	Mean
2.	Variance
3.	Skewness
4.	Kurtosis
5.	Entropy
6.	Energy

In this paper, the value obtained from our work for different type of image is given as follows:

Table 3.1.(b) Intensity histogram features and their values

Image Type	Features					
	Mean	**Variance**	**Skewness**	**Kurtosis**	**Entropy**	**Energy**
normal	7.2534	1.6909	-1.4745	7.8097	0.2504	1.5152
malignant	6.8175	4.0981	-1.3672	4.7321	0.1904	1.5555
benign	5.6279	3.1830	-1.4769	4.9638	0.2682	1.5690

3.2 GLCM Features

It is a statistical method that considers the spatial relationship of pixels is the gray-level co-occurrence matrix (GLCM), also known as the gray-level spatial dependence matrix. By default, the spatial relationship is defined as the pixel of interest and the pixel to its immediate right (horizontally adjacent), but you can specify other spatial relationships between the two pixels. Each element (I, J) in the resultant GLCM is simply the sum of the number of times that the pixel with value *I* occurred in the specified spatial relationship to a pixel with value *J* in the input image.

The Following GLCM features were extracted in our research work:

Autocorrelation, Contrast, Correlation, Cluster Prominence, Cluster Shade, Dissimilarity Energy, Entropy, Homogeneity, Maximum probability, Sum of squares, Sum average, Sum variance, Sum entropy, Difference variance, Difference entropy, information measure of correlation, Inverse difference normalized.

The value obtained for the above features from our work for a typical image is given in the following table 3.2

Table 3.2: GLCM Features and values Extracted from Mammogram Image

Feature No	Feature Name	Feature Values
1	Autocorrelation	44.1530
2	Contrast	1.8927
3	Correlation	0.1592
4	Cluster Prominence	37.6933
5	Cluster Shade	4.2662
6	Dissimilarity	0.8877
7	Energy	0.1033
8	Entropy	2.6098
9	Homogeneity	0.6645
10	Maximum probability	0.6411
11	Sum of squares	0.1973,
12	Sum average	44.9329
13	Sum variance	13.2626
14	Sum entropy	133.5676
15	Difference variance	1.8188
16	Difference entropy	1.8927
17	Information measure of correlation	1.2145
18	Inverse difference normalized	0.2863

4. FEATURE SELECTION

Feature selection helps to reduce the feature space which improves the prediction accuracy and minimizes the computation time. This is achieved by removing irrelevant, redundant and noisy features .i.e., it selects the subset of features that can achieve the best performance in terms of accuracy and computation time. It performs the Dimensionality reduction.

Features are generally selected by search procedures. A number of search procedures have been proposed. Popularly used feature selection algorithms are Sequential forward Selection, Sequential Backward selection, Genetic Algorithm and Particle Swarm Optimization, Branch and Bound feature optimization. In this work a combined approach of Fast Branch and Bound technique for optimal feature selection algorithm and Genetic Algorithm is proposed to select the optimal features. The selected optimal features are considered for classification. Till now no attempts have been made to hybrid the different feature selection algorithm to extract the feature from mammogram. Especially branch and bound techniques has been fully exploited to extract the feature from mammogram which is one of the best techniques to optimize the features among many features. We have attempted to optimize the feature of GLSM by fast branch and bound technique as well as Genetic algorithm to extract most relevant features for our classification work.

4.1 Genetic algorithms

As an optimization technique, Genetic Algorithms simultaneously examine and manipulate a set of possible solutions. The GA starts with several alternative solutions to the optimization problem, which are considered as individuals in a population. These solutions are coded as binary strings, called chromosomes. The initial population is constructed randomly. These individuals are evaluated, using the partitioning-specific fitness function. The GA then uses these individuals to produce a new generation of hopefully better solutions. In each generation, two of the individuals are selected probabilistically as parents, with the selection probability proportional to their fitness. Crossover is performed on these individuals to generate two new individuals, called offspring, by exchanging parts of their structure. Thus each offspring inherits a combination of features from both parents. The next step is mutation. An incremental change is made to each member of the population, with a small probability. This ensures that the GA can explore new features that may not be in the population yet. It makes the entire search space reachable, despite the finite population size. Roulette Wheel parent selection method which is conceptually the simplest stochastic selection technique. Our generation replacement technique is based on replacing the most inferior member in a population by new offspring

4.2 The Fast Branch & Bound Algorithm

Consider a problem of selecting d features from an initial set of D measurements using objective function J as a criterion of effectiveness. The Branch & Bound approach aims to solve this search problem by making use of the monotonicity property of certain feature selection criterion function.

Let $\mathcal{X}_j$ be the set of features obtained by removing j features $y_1, y_2, \cdots, y_j$ from the set Y of all D features, i.e.,

$$\mathcal{X}_j = Y \setminus \{y_l, y_2, \cdots, y_j\}: \quad (l)$$

The monotonicity condition assumes that, for feature subsets

$\mathcal{X}_l, \mathcal{X}_2, \cdots, \mathcal{X}_j$, where $\mathcal{X}_l$ subset $\mathcal{X}_2$ subset ......$\mathcal{X}_j$,
the criterion function **J**
fulfills

$$\mathbf{J}(\mathcal{X}_l) >= \mathbf{J}(\mathcal{X}_2) >= \cdots >= \mathbf{J}(\mathcal{X}_j): \quad (2)$$

The monotonicity property helps to identify parts of the search space, which cannot possibly contain the optimal solution to the feature selection problem.

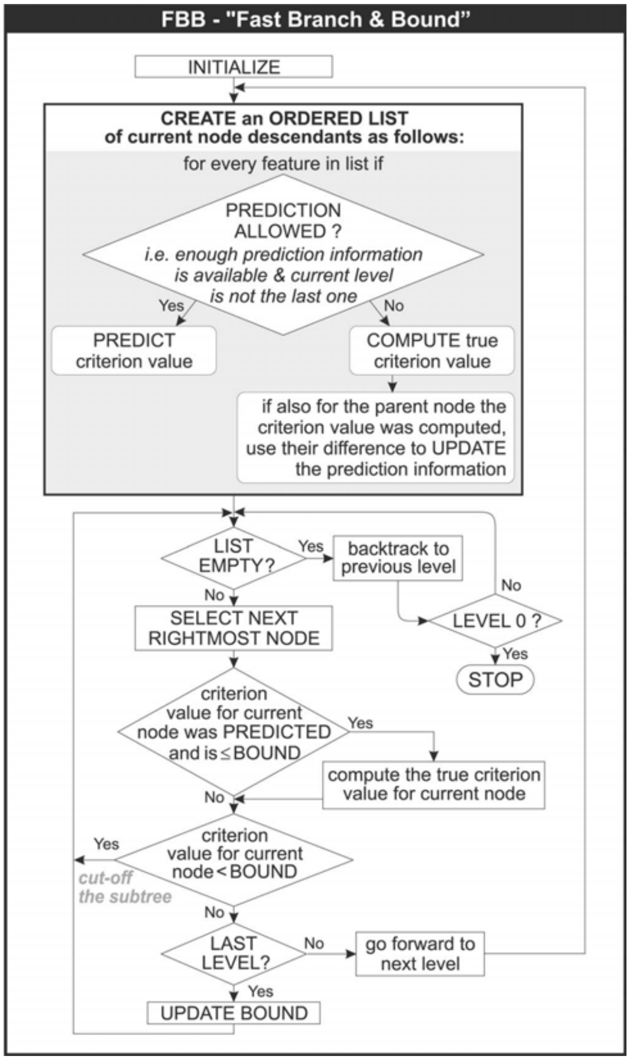
FBB - "Fast Branch & Bound"
INITIALIZE
CREATE an ORDERED LIST
of current node descendants as follows:
for every feature in list if
PREDICTION ALLOWED ?
i.e. enough prediction information is available & current level is not the last one
Yes
No
PREDICT criterion value
COMPUTE true criterion value
if also for the parent node the criterion value was computed, use their difference to UPDATE the prediction information
LIST EMPTY?
Yes
backtrack to previous level
No
LEVEL 0 ?
Yes
STOP
No
SELECT NEXT RIGHTMOST NODE
criterion value for current node was PREDICTED and is ≤ BOUND
Yes
compute the true criterion value for current node
No
criterion value for current node < BOUND
Yes
cut-off the subtree
No
LAST LEVEL?
No
go forward to next level
Yes
UPDATE BOUND

Fig. 3 Fast Branch and Bound Algorithm

4.3 Proposed Hybrid Approach Algorithm:

1. Extract N number of features A1, A2, A3..AN from ROI Of the preprocessed Image
2. Apply Genetic algorithm to select the optimal set containing n1 number of features where n1<N
3. Apply Branch and Bound feature optimization [23] search to select the best subset containing n2 number of features n2 where n2<N
4. Find the Union of n1 features and n2 features as n features
5. Use the n features where n<N for Classification.

The selected features using GA method are tabulated as follows:

Table 4.1(a): Feature selected By GA method

S.no	Features
1	Cluster prominence
2	Energy
3	Information measure of correlation
4	Inverse difference Normalized
5	Skewness
6	Kurtosis

The selected features using Branch and Bound method are listed in the following table

Table 4.1(b): Feature selected By Branch and Bound Method for feature subset selection.

S.no	Feature
1	Energy
2	Mean
3	Contrast
4	Variance
5	Information measure of correlation
6	Skewness

By applying the proposed algorithm, it will produce a feature set contain best set of features which is less than the original set. This two different methods will be providing a better and concrete relevant feature selection from 18 nos. of features to minimize the classification time and error and productive results in conjunction with better accuracy positively. The genetic algorithm has shown a better result to maximum extends and the actual fast branch and bound algorithm has shown more accurate relevant features to our classification. It uses a tree structure and use bhatacharya distance metric to bound the variable features to optimize the selection. The union of features from the two methods is given in the below table

Table 4.1 (c): Feature selected by proposed Hybrid method

S.no.	Features
1	Cluster prominence
2	Energy
3	Information measure of correlation
4	Inverse difference Normalized
5	Skewness
6	Kurtosis
7	Contrast
8	Mean
9	Variance

4.4 Classification

A decision tree is typically; evaluated by predictive accuracy that considers all errors equally. However, the predictive accuracy might not be appropriate when the data is imbalanced and/or the costs of different errors vary markedly. As an example, consider the classification of pixels in mammogram images as possibly cancerous (Woods et al., 1993; Chawla et al., 2002). A typical mammography data set might contain 98% normal pixels and 2% abnormal pixels. A simple default strategy of guessing the majority class would give a predictive accuracy of 98%. Ideally, a fairly high rate of correct cancerous predictions is required, while allowing for a small to moderate error rate in the majority class. It is more costly to predict a cancerous case as non-cancerous, than otherwise.

Moreover, distribution/cost sensitive applications can require a ranking or a probabilistic estimate of the instances. For instance, revisiting our mammography data example, a probabilistic estimate or ranking of cancerous cases can be decisive for the practitioner. The cost of further tests can be decreased by thresholding the patients at a particular rank. Secondly, probabilistic estimates can allow one to threshold ranking for class membership at values < 0:5. Hence, the classes assigned at the leaves of the decision trees have to be appropriately converted to probabilistic estimates (Provost & Domingos, 2003; Zadrozny & Elkan, 2001). This brings us to another question: What is the right probabilistic estimate for imbalanced data sets? We attempt to answer the questions raised in the preceding discussion using C4.5 release 8 decision trees as our classifier.

4.4.1 Probabilistic C4.5

Typically, C4.5 assigns the frequency of the correct counts at the leaf as the probabilistic estimate. For notational purposes, TP is the number of true positives at the leaf, FP is the number of false positives, and C is the number of classes in the data set. Thus, the frequency based probabilistic estimate can be written as: Pleaf = TP=(TP + FP)

However, simply using the frequency of the correct counts (of classes) at a leaf might not give sound probabilistic estimates (Provost & Domingos, 2003;Zadrozny & Elkan, 2001). A (small) leaf can potentially give optimistic estimates for classi_cation purposes. For instance, the frequency-based estimate will give the same weights to leaves with the following (TP; FP) distributions: (5; 0) and (50; 0). The relative coverage of the leaves and the original class distribution is not taken into consideration. Given the evidence, a probabilistic estimate of 1 for the (5; 0) leaf is not very sound. Smoothing the frequency-based estimates can mitigate the aforementioned problem

(Provost & Domingos, 2003). One way of smoothing those probabilities is using the Laplace estimate, which can be written as follows:

PLaplace = (TP + 1)=(TP + FP + C)

Again considering the two cases of TP =5 and TP = 50, the Laplace estimates are 0.86 and 0.98, respectively, which are more reliable given the evidence. However, Laplace estimates might not be very appropriate for highly imbalanced data sets (Zadrozny & Elkan, 2001). In that scenario, it could be useful to incorporate the prior of positive class to smooth the probabilities so that the estimates are shifted towards the minority class base rate (b). The m-estimate (Cussents, 1993) can be used as follows (Zadrozny & Elkan, 2001):

Pm = (TP + bm)=(TP + FP +m) where b is the base rate or the prior of positive class, and m is the parameter for controlling the shift to wards b. Zadrozny and Elkan (2001) suggest using m, given b, such that bm = 10.

4.4.2 Algorithm

C4.5 builds decision trees from a set of training data in the same way as ID3, using the concept of information entropy. The training data is a set $S = s_1,s_2,...$ of already classified samples. Each sample $s_i = x_1,x_2,...$ is a vector where $x_1,x_2,...$ represent attributes or features of the sample. The training data is augmented with a vector $C = c_1,c_2,...$ where $c_1,c_2,...$ represent the class to which each sample belongs.

At each node of the tree, C4.5 chooses one attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. Its criterion is the normalized information gain (difference in entropy) that results from choosing an attribute for splitting the data. The attribute with the highest normalized information gain is chosen to make the decision. The C4.5 algorithm then recurs on the smaller sublists.

This algorithm has a few base cases.

- All the samples in the list belong to the same class. When this happens, it simply creates a leaf node for the decision tree saying to choose that class.
- None of the features provide any information gain. In this case, C4.5 creates a decision node higher up the tree using the expected value of the class.
- Instance of previously-unseen class encountered. Again, C4.5 creates a decision node higher up the tree using the expected value.

In pseudocode, the general algorithm for building decision trees is:

```
1. Check for base cases

2. For each attribute a

       (i) find the normalized information gain
           from splitting on a

3. Let a_best be the attribute with the highest
   normalized information gain

4. Create a decision node that splits on a_best

5. Recur on the sublists obtained by splitting on
   a_best, and add those nodes as children of node
```

J48 is an open source Java implementation of the C4.5 algorithm in the weka data mining tool.

The selected features are used for classification. For classification of samples, we have employed the freely available Machine Learning package, WEKA [4] to train our data set using J48 decision tree method. Out of 300 images in the dataset, 208 were used for training and the remaining 92 for testing purposes

5. EXPERIMENTAL RESULTS

In this paper we used J48 classifier, a decision tree classifier based on C4.5, from WEKA [4] to train and test the features. The average accuracy is 95%. We have used the precision and recall measures as the evaluation metric for mammogram classification. Precision is the fraction of the number of true positive predictions divided by the total number of true positives in the set. Recall is the total number of predictions divided by the total number of true positives in the set.

The testing results using the selected features are given as shown in the table 5.1

Table 5.1 Results obtained by proposed method

Normal	100%
Malignant	89.6%
Benign	100%

The confusion matrix has been obtained from the testing part .In this case for example out of 97 actual malignant images 10 images was classified as normal. In case of benign and normal all images are correctly. The confusion matrix is given below (Table 5.2)

Table5.2: Confusion matrix

Actual	Predicted class		
	Benign	Malignant	Normal
Benign	104	0	0
Malignant	97	87	10
Normal	99	0	99

The following graph shows the comparative analysis of our method and various other methods.

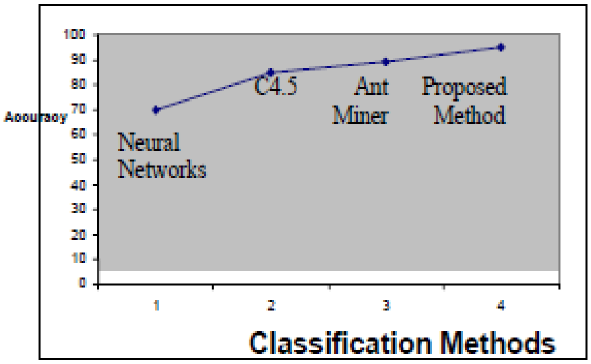

Fig. 4 Performance of the Classifier

6. CONCLUSION

Mammography is one of the best methods in breast cancer detection, but in some cases radiologists face difficulty in directing the tumors. The methods like one presented in this paper could assist the medical staff and improve the accuracy of detection. Our method can reduce the computation cost of mammogram image analysis and can be applied to other image analysis applications. The algorithm uses simple statistical techniques in collaboration to develop a novel feature selection technique for medical image analysis. The value of this technique is that it not only tackles the measurement problem but also provides a visualization of the relation among features. In addition to ease of use, this approach effectively addresses the feature redundancy problem. The method proposed has been proven that it is easier and it requires less computing time than existing methods.

7. REFERENCES

[1]. Majid AS, de Paredes ES, Doherty RD, Sharma N Salvador X. “Missed breast carcinoma: pitfalls and Pearls”. Radiographics 23(2003)881-895.

[2]. Osmar R. Zaïane,M-L. Antonie, A. Coman “Mammography Classification by Association Rulebased Classifier,” MDM/KDD2002 International Workshop on Multimedia Data Mining with (ACM SIGKDD 2002, Edmonton, Alberta, Canada, 17-19 July 2002,), pp.62-69.

[3]. Xie Xuanyang, Gong Yuchang, Wan Shouhong, Li Xi ,”Computer Aided Detection of SARS Based on Radiographs Data Mining “, Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference Shanghai, China, September 1-4, 2005

[4]. Shuyan Wang, Mingquan Zhou and Guohua Geng, “Application of Fuzzy Cluster analysis for Medical Image Data Mining” Proceedings of the IEEE International Conference on Mechatronics & Automation Niagara Falls, Canada • July 2005.

[5]. R.Jensen, Qiang Shen, “Semantics Preserving Dimensionality Reduction: Rough and Fuzzy-Rough Based Approaches”, IEEE Transactions on Knowledge and Data Engineering, 16 (2004) 1457-1471.

[6]. Walid Erray, and Hakim Hacid, “A New Cost Sensitive Decision Tree Method Application for Mammograms Classification” IJCSNS International Journal of Computer Science and Network Security, 6 (2006) No.11.

[7]. Ying Liu, Dengsheng Zhang, Guojun Lu, “Regionbased image retrieval with high-level semantics using decision tree learning”, Pattern Recognition, 41 (2008) 2554 – 2570

[8]. Kemal Polat , Salih Gu¨nes, A novel hybrid intelligent method based on C4.5 decision tree classifier and one-against-all approach for multi-class classification problems, Expert Systems with Applications,

[9]. R. S. Parpinelli, H. S. Lopesand A. A. Freitas, “Data Mining with an Ant Colony Optimization Algorithm”, IEEE Transactions on Evolutionary Computing, 6 (2002) No. 4 : 321-332

[10].Bo Liu, Hussein A. Abbass, Bob McKay. “Classification Rule Discovery with Ant Colony Optimization”, IEEEkomputational Intelligence Bulletin 3 (February 2004) No.1.

[11]. P. Jaganathan, K.Thangavel , A. Pethalakshmi, M. Karnan, “Classification Rule Discovery with Ant Colony Optimization and Improved Quick Reduct Algorithm”, IAENG International Journal ofComputer Science, 33:1, IJCS_33_1_9

[12].K.Thangavel, M.Karnan, “Computer Aided Diagnosis in Digital Mammograms: Detection of Microcalcifications by Metaheuristic Algorithms.” ICGST-GVIP Journal, 5 (July 2005) Issue (7).

[13]. Jiang J, Yao B, Wason AM, A genetic algorithm design for microcalcification detection and classification in digital mammograms , Comput Med Imaging Graph 2007 Jan;31(1):49-61.

[14].Gonzalo V. Sánchez-Ferrero,Juan Ignacio Arribas, “A Statistical-Genetic Algorithm to Select the Most Significant Features in Mammograms” , Springer Berlin Springer Berlin / Heidelberg,(August 18, 200) pp 189-196

[15]. Haralick RM, Shanmugam K, Dinstein I. “Textural features for image classification” IEEE Transactions On Systems, Man, and Cybernetics, 3 (1973) pp 610–621.

[16]. Dougherty J, Kohavi R, Sahami M. “Supervised and unsupervised discretization of continuous features”. In: Proceedings of the 12th international conference on machine learning.San Francisco:Morgan Kaufmann; (1995) pp 194–202.

[17] D.Brazokovic and M.Nescovic ., “Mammogram screening using multisolution based image segmentation”, International journal of pattern recognition and Artificial Intelligence, 7(6):1437-1460,1993.

[18] C.Chen and G.Lee, “Image segmentation using multitiresolution wavelet analysis and Expectation Maximum(EM) algorithm for mammography” , International Journal of Imaging System and Technology, 8(5):491-504,1997.

[19] I.Christiyanni et al ., “Fast detection of masses in computer aided mammography”, IEEE Signal processing Magazine, Pages:54-64,2000.

[20] Holmes, G., Donkin, A., Witten, I.H.: “WEKA: a machine learning workbench.” In: Proceedings Second Australia and New Zealand Conference on Intelligent Information Systems, Brisbane, Australia, pp. 357-361 (1994).

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

[22] T.Wang and N.Karayaiannis, “Detection of microcalcification in digital mammograms using wavelets”, IEEE Trans. Medical Imaging, 17(4):498-509,1998

[23] S.Petre,P.Pavel, K.Josef “Fast Branch & Bound Algorithms for Optimal Feature Selection” IEEE Trans on Pattern Analysis and Machine Intelligence Vol. 26, No. 7

[24] Blake, C., & Merz, C. (1998). UCI Repository of MachineLearningDatabaseshttp://www.ics.uci.edu/_mlearn/_MLRepository.html.Department of Information and Computer Sciences,University of California, Irvine.

[25] Bradley, A. P. (1997). The Use of the Area Under theROC Curve in the Evaluation of Machine Learning Algorithms. Pattern Recognition, 30(6), 1145{1159.

[26] Chawla, N., Hall, L., K.W., B., & Kegelmeyer, W. (2002). SMOTE: Synthetic Minority Over sampling Technique. Journal of Artificial Intelligence Research, 16, 321{357.

[27] Cohen, W. (1995). Learning to Classify English Textwith ILP Methods. Proceedings of the 5th International Workshop on Inductive Logic Programming (pp. 3{24). Department of Computer Science, Katholieke Universiteit Leuven.

[28] Cost, S., & Salzberg, S. (1993). A Weighted Nearest Neighbor Algorithm for Learning with Symbolic Features. Machine Learning, 10, 57{78}.

[29] Cussents, J. (1993). Bayes and pseudo-bayes estimates of conditional probabilities and their reliabil-ities. Proceedings of European Conference on Machine Learning.

[30] Dumais, S., Platt, J., Heckerman, D., & Sahami, M.(1998).Inductive Learning Algorithms and Representations for Text Categorization. Proceedings of the Seventh International Conference on Information and Knowledge Management. (pp. 148{155).

[31] Ezawa, K., J., Singh, M., & Norton, S., W. (1996). Learning Goal Oriented Bayesian Networks for Telecommunications Risk Management. Proceedings of the International Conference on Machine Learning, ICML-96 (pp. 139{147). Bari, Italy: Morgan

[32] Kau_man.Fawcett, T., & Provost, F. (1996). Combining Data Mining and Machine Learning for Elective User Prole. Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining(pp. 8{13). Portland, OR: AAAI.

[34] Hand, D. (1997). Construction and assessment of classification rules. Chichester: John Wiley and Sons.
Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. Intelligent Data Analysis, 6.

[37] Kubat, M., Holte, R., & Matwin, S. (1998). Machine Learning for the Detection of Oil Spills in Satellite Radar Images. Machine Learning, 30, 195{215.

[38] Lewis, D., & Ringuette, M. (1994). A Comparison of Two Learning Algorithms for Text Categorization. Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval (pp. 81{93).

[39] Mladeni_c, D., & Grobelnik, M. (1999). Feature Selection for Unbalanced Class Distribution and Naive Bayes. Proceedings of the 16th International Conference on Machine Learning. (pp. 258{267).

[40] MorganKaufmann. Provost, F., & Domingos, P. (2003). Tree induction for probability-based rankings. Machine Learning,52(3).

[41] Quinlan, J. (1992). C4.5: Programs for Machine Learning. San Mateo, CA: Morgan Kaufmann Stan_ll, C., & Waltz, D. (1986). Toward Memory-based Reasoning. Communications of the ACM, 29, 1213{1228.

[42] Swets, J. (1988). Measuring the Accuracy of Diagnostic Systems. Science, 240, 1285{1293.

[43] Weiss, G. (1995). Learning with rare cases and small disjuncts. Proceedings of the Twelfth International Conference on Machine Learning.

[44] Woods, K., Doss, C., Bowyer, K., Solka, J., Priebe, C.,& Kegelmeyer, P. (1993). Comparative Evaluation of Pattern Recognition Techniques for Detection of Microcalci_cations in Mammography. International Journal of Pattern Recognition and Arti_cial Intelligence, 7(6), 1417{1436.

[45] Zadrozny, B., & Elkan, C. (2001). Learning and making decisions when costs and probabilities are both unknown. Proceedings of the Seventh International Conference on Knowledge Discovery and Data Mining.