

# Feature Extraction and Rule Classification Algorithm of Digital Mammography based on Rough Set Theory

About Ella Hassanien

Jafar M. H. Ali.

Kuwait University, Faculty of Administrative Science,  
Quantitative Methods and Information Systems Department  
P.O..Box 5969 Safat, code no. 13060 Kuwait

---

**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 an efficient classification algorithm in digital mammograms in the context of rough set theory. Feature extractions acquired in this work are derived from the gray-level co-occurrence matrix. The features are extracted, normalized and then the rough set dependency rules are generated directly from the real value attribute vector. Then the classification is built and the quadratic distance function used to determines similarity between a query and database image. The experimental results show that the proposed algorithm performs well reaching over 85% in accuracy.*

**Keywords:** *rough sets, mammograms, classification, feature extraction, rule generation, similarity measure, gray-level co-occurrence matrices*

## 1. Introduction

Breast cancer is a deadly disease that adversely affects the lives of far too many people, primarily women. According to the National Cancer Institute [5], each year about 180,000 women in the United States develop breast cancer, and about 48,000 lose their lives to this disease. It is also reported that a woman's lifetime risk of developing breast cancer is one in eight. Currently, digital mammography [11,13,15,16] is one of the most promising cancer control strategies since the cause of breast cancer is still unknown. Mammography is a specific type of imaging that uses a low-dose x-ray system and high-contrast, high-resolution film for examination of the breasts. Most medical experts agree that successful treatment of breast cancer often is linked to early diagnosis. Mammography plays a central part in early detection of breast cancers because it can show changes in the breast up to two years before a patient or physician can feel them. Classification is a form of medical data analysis, which can be used to extract models describing important data classes or to predict future data trends. In other words, classification is to identify essential features of different classes based on a training set and then classify new instances into the appropriate classes [8,12]. Classification of digital mammogram of breast lesions as malignant or benign must be based on information present in the mammogram images. In additional, clinical data such as patient age can also be used. A common approach to this task is to extract information (called features) from mammograms and then build a classifier model based on rough set theory to make the malignant versus benign assessment.

In this paper, we introduce an efficient classification approach based on the context of rough set theory which applied on five statistical extracted features from the digital mammograms. Rough set concept was introduced by Polish logician, Professor Zdzisław Pawlak in early eighties [17]. This theory become very popular among scientists around the world and the rough set is now one of the most developing

intelligent data analysis. Rough sets data analysis was used for the discovery of data dependencies, data reduction, approximate set classification, and rule induction from databases. The generated rules represent the underlying semantic content of the images in the database. A classification mechanism is developed by which the images are classified according to the generated rules. The analysis of medical, biological and health-related data has been one of the successful applications of rough sets. In some cases, the rough set analysis has given medical doctors new insight into their practice and training.

This paper is organized as follows. Section 2 depicts the general classification process. Section 3 gives a brief description of the image pre-processing phase. Feature extraction based on Gray-level Co-occurrence Matrix is presented in Section 4. In section 5, the fundamental of rough set theory is introduced. The rule generation and rule classification algorithm is discussed in Section 6. Experimental results are given and discussed in section 7. The paper is concluded in section 8.

## **2. Image Classification Process**

Classification is a function that classifies a data item into one of several predefined classes. In other words, classification is the process that establishes classes with attributes from a set of instances (called training sets). The class of an instance must be one from a finite set of possible, pre-determined class values, while attributes of the instance are descriptions of the instance potentially affecting its class. A training set is a set of tuples defined on a set of multiple attributes (also, called features). Each tuple has a known class label, which is called a target class, associated with it [14]. The aim of classification process is to induce a classifier that can be used to classify future data items whose classification is unknown. The classification is based on a well-defined set of classes and a training set of pre-classified examples (tuples). The knowledge produced during the classification process can be extracted and represented in the form of rules. These classification rules allow one to develop a profile of items belonging to a particular group according to their common attributes. This profile can then be used to classify new data items that are added to the database. The accuracy of a classifier on a given test set is the percentage of test set samples that are correctly classified by the model. If the accuracy of the classifier is considered acceptable, the classifier can be used to classify future data tuples or objects for which the class label is not known. Figure (1) illustrates the classification scheme. The architecture for image classification under the rough set theory framework [2] is composed by four fundamental building phases: pre-processing, feature extraction, rule generation and similarity function.

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. After enhancing the images, feature relevant to classification are extracted and represented in a database. Given a feature representation for each database image, retrieval consists of extracting a set of feature vectors from the queried image and relying on a similarity function to evaluate which feature representation best explains those features. In feature representation, a scheme is needed to convert an attribute vector to be read by the rough set rule generator. Rule generation forms the core of the rough set framework. The prediction of new objects is dependent on the type of rules generated and some strategies will work well for prediction. Rules generated can be

used to predict values of decision attributes (classify) for new object. The rules of the query specification are compared with rules of the image database to determine which images match correctly (similar) with the given rules. The matching process is based on similarity measure between query image and images in database.

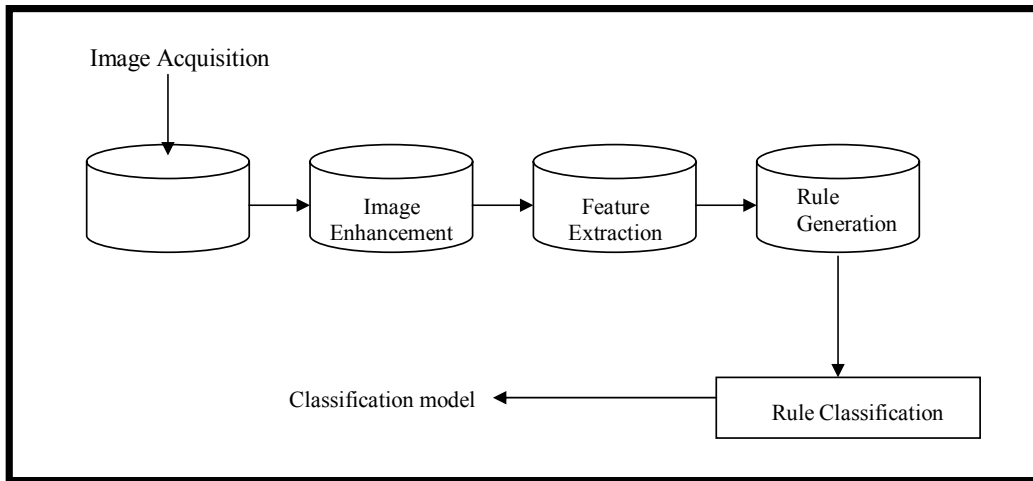


Figure 1: classification scheme

### 3. Pre-Processing phase

Diagnosing cancer tissues using digital mammograms is a time consuming task even for highly skilled radiologists because mammograms are low contrast, noisy images. Therefore, in digital mammogram there is a need for enhancing imaging before a reasonable feature extraction can be achieved. Image enhancement in medical computing is the use of computers to make an image clearer [3]. This may be to aid interpretation by humans or computers. Types of image enhancement include, noise reduction, edge enhancement and contrast enhancement.

Contrast enhancement is useful when an area of the image that is of particular importance has only subtle changes in pixel intensity. In these cases it may be difficult for the human eye to make out the structures clearly, especially if the image is being displayed on a low quality screen, or limited contrast like mammograms image. Applying the contrast enhancement filter will improve the readability of areas with subtle changes in contrast but will also destroy areas of the image where the intensity of the pixels is outside the range of intensities being enhanced. In this work, we applied pre-processing techniques based on mathematical morphology theory to enhancing the contrast of the original mammogram image to help in feature extraction process [3, 6].

### 4. Texture Feature Extraction

Texture is one of the most important defining characteristics of an image. It is characterized by the spatial distribution of gray levels in a neighborhood [9]. In order to capture the spatial dependence of gray-level values which contribute to the perception of texture, a two-dimensional dependence texture analysis matrix is discussed for texture

consideration. Since texture shows its characteristics by both each pixel and pixel values. There are many approaches using for texture classification. The gray-level co-occurrence matrix seems to be a well-know statistical technique for feature extraction [4, 14].

#### 4.1 Gray-level Co-occurrence Matrix

Co-occurrence matrix, the second-order histogram, is the two dimensional matrix of joint probabilities  $P_{d,r}(i, j)$  between pairs of pixels, separated by a distance  $d$  in a given direction  $r$ . Haralick et al., [9] define 14 texture features derived from the co-occurrence matrix. In this paper, five features were selected for further study; maximum probability, contrast, inverse different moment, angular second moment, and entropy. Table (1) gives the description of these selected features. Where  $P(i, j)$  refers to the normalized entry of the co-occurrence matrices. That is  $P(i, j) = P_d(i, j) / R$  where  $R$  is the total number of pixel pairs  $(i, j)$ . For a displacement vector  $d = (dx, dy)$  and image of size  $N \times M$   $R$  is given by  $(N - dx)(M - dy)$ .

Feature	Equation	Description
Maximum probability	$\max_{i,j} \sum P(i, j)$	This is simply the largest entry in the matrix, and corresponds to the strongest response.
Contrast	$\sum_{i,j} (i - j)^2 P(i, j)$	A measure of the image contrast or the amount of local variations present in an image.
Inverse different moment	$\sum_{i,j} \frac{1}{1 + (i - j)^2} P(i, j)$	This descriptor has large values in cases where the largest elements in P are along the principal diagonal.
Angular second moment	$\sum_{i,j} P(i, j)^2$	A measure of the homogeneity of an image. Hence it is a suitable measure for detection of disorders in textures. For homogeneous textures value of angular second moment turns out to be small compared to non-homogeneous ones.
Entropy	$\sum_{i,j} P(i, j) \log(P(i, j))$	Entropy is a measure of information content. It measures the randomness of intensity distribution.

Table 1: texture features extracted from Co-occurrence Matrix

#### 5. Rough Set Theory: Theoretical Background

Let us present here some preliminaries of rough set theory, which are relevant to this work. For details one may refer to [7, 17, and 18].

## 5.1 Information Systems

Knowledge representation in rough sets is done via information systems, which are a tabular form of an OBJECT  $\rightarrow$  ATTRIBUTE VALUE relationship. More precisely, an information system,  $\Gamma = \langle U, \Omega, V_q, f_q \rangle_{q \in \Omega}$ , where

- $U$  is a finite set of objects,  $U = \{x_1, x_2, x_3, \dots, x_n\}$
- $\Omega$  is a finite set of attributes (features), the attributes in  $\Omega$  are further classified into disjoint condition attributes  $A$  and decision attributes  $D$ ,  $\Omega = A \cup D$
- For each  $q \in \Omega$ ,
  - $V_q$  is a set of attribute values for  $q$ ,
  - Each  $f_q: U \rightarrow V_q$  is an information function which assigns particular values from domains of attributes to objects such that  $f_q(x_i) \in V_q$  for all  $x_i \in U$  and  $q \in \Omega$ .

With respect to a given  $q$ , the functions partitions the universe into a set of pairwise disjoint subsets of  $U$ :

$$R_q = \{x : x \in U \wedge f(x, q) = f(x_0, q) \forall x_0 \in U\} \quad (1)$$

Assume a subset of the set of attributes,  $P \subseteq A$ . Two samples  $x$  and  $y$  in  $U$  are indiscernible with respect to  $P$  if and only if  $f(x, q) = f(y, q) \forall q \in P$ . The indiscernibility relation for all  $P \subseteq A$  is written as  $IND(P)$ .  $U / IND(P)$  is used to denote the partition of  $U$  given  $IND(P)$  and is calculated as follows:

$$U / IND(P) = \otimes \{q \in P : U / IND(P)(\{q\})\}, \quad (2)$$

$$A \otimes B = \{X \cap Y : \forall q \in A, \forall Y \in B, X \cap Y \neq \{\}\}. \quad (3)$$

## 5.2 Approximation Spaces

A rough set approximates traditional sets using a pair of sets named the lower and upper approximation of the set. The lower and upper approximations of a set  $P \subseteq U$ , are defined by equations (4) and (5), respectively.

$$\underline{PY} = \bigcup \{X : X \in U / IND(P), X \subseteq Y\} \quad (4)$$

$$\overline{PY} = \bigcup \{X : X \in U / IND(P), X \cap Y \neq \{\}\} \quad (5)$$

Assuming  $P$  and  $Q$  are equivalence relations in  $U$ , the important concept positive region  $POS_p(Q)$  is defined as:

$$POS_p(Q) = \bigcup_{X \in Q} \underline{PX} \quad (6)$$

A positive region contains all patterns in  $U$  that can be classified in attribute set  $Q$  using the information in attribute set  $P$ .

## 6. Building the Classifier

The goal of classification is to assign a new object to a class from a given set of classes based on the attribute values of this object. To classify objects, which has never been seen before, rules generated from a training set will be used. These rules represent the actual classifier. This classifier 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 classification [1]. This section describes how the classification system is built and how a new pattern can be classified using this system. Given an object to classify, the features discussed in section (4) are extracted. The features in the object would yield a list of applicable rules. Then the applicable rules are grouped by class in their consequent part and the groups are ordered by the sum of rules confidences, the ordered groups would indicate the most significant class that should be attached to the object to be classified. Figure (2) illustrates the rule classification algorithm. We use the already generated rules to classify new objects. Given a new image its feature vector is first extracted and then the attribute vector is computed. The nearest matching rule is determined as the one whose condition part differs from the attribute vector of re-image by the minimum number of attributes. When there is more than one matching rule, we use a voting mechanism to choose the decision value. Every matched rule contributes votes to its decision value, which are equal to the  $t$  times number of objects matched by the rule. The votes are added and the decision with the largest number of votes is chosen as the correct class. Here  $t$  represents the validity of the rule.

---

### Algorithm: Classification of a new object

**Input:** A new image to be classified, the attribute vector of the new image, and the set of rules

**Output:** The final classification

**Processing: Begin**

For each rule in Rule set Do

If match (rule, new object) Then  
 Measure = |Objects|,  $K \rightarrow$  |Classes|;  
 For  $i=1$  to  $K$  Do

Collect the set of objects defining the concept  $X^i$

Extract  $Mrule(X^i, u^t) = \{r \in Rule\}$

For any rule  $r \in Mrule(X^i, u^t)$  Do

$T = Match^A(r) \cap X^i$  and  $LL = LL \cup T$ ;

Strength =  $Card(LL) / Card(X^i)$

Vote = Measure \* Strength

Give Vote(Class(Rule), Vote)

Return Class with highest Vote

End

---

Figure 2: rule classification algorithm

## 6.1 Similarity Measure

To select the top  $N$  matches to a query image, the database is ranked based on the quadratic distances between the query model and models of each candidate images. The quadratic distance between two  $m$ -dimensional histograms  $h_e$  and  $h_p$  is defined as:

$$d_{Quadratic}^2(e, p) = (h_e - h_p)^T A (h_e - h_p) \quad (8)$$

Where  $A$  is a matrix of similarity weights,  $A = [a_{ij}]$ ,  $0 \leq a_{ij} < 1$ ,  $a_{ii} = 1$  and

$$a_{i,j} = \begin{cases} 1 - d_{i,j} / d_{\max} & d_{i,j} \leq T_d \\ 0 & d_{i,j} > T_d \end{cases} \quad (9)$$

$d_{ij}$  is the Euclidean distance between colors  $i$  and  $j$ , and  $d_{\max}$  is the greater distance between gray colors. That is, coefficients  $a_{ij}$  for two gray colors are defined by:

$$m_0 = (h_e, s_e, v_e) \text{ and } m_1 = (h_p, s_p, v_p).$$

## 7. Results and Discussion

### 7.1 Data sets

The data sets that we used in this work were taken from the Mammography Image Analysis Society (MIAS) [10]. It contains 320 images, which belong to three normal categories: normal, benign and malign. There are 206 normal images, 63 benign and 51 malign, which are considered abnormal. In addition, the abnormal case are further divided in six categories: microcalcification, circumscribed masses, speculated masses, ill-defined masses, architectural distortion and asymmetry. All the images also include the locations of any abnormalities that may be present. We divide the 320 samples of mammogram images into 10 equal size folders, such that a single folder is used for testing the model that has been developed from the remaining nine sets. The evaluation statistics for each method is then assessed as an average of 10 experiments.

### 7.2 Visual features

The query was performed by providing a query image from a data set and the selected five texture features: maximum probability, contrast, inverse difference moment, angular second moment and entropy calculated from each occurrence matrix and their values are saved in the feature vector of the corresponding image. Then the rules will be generated and ordered. The similarity between images is estimated by summing up the distance between corresponding features in their feature vectors. Images having feature vectors closest to feature vector of the query image are returned as best matches. The results were then numerically sorted and the best 12 images were displayed along with the query image.

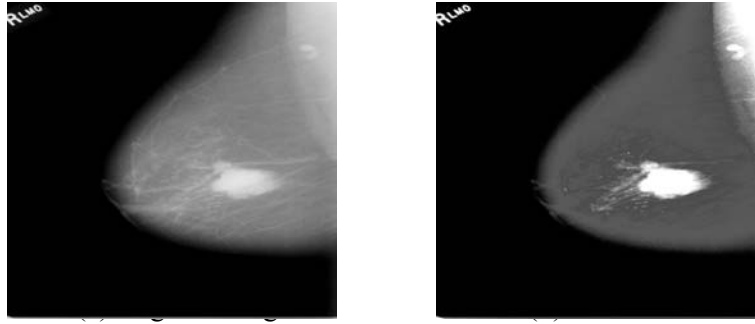


Figure 3: Pre-processing phase on an example image

### 7.3 Performance Evaluation for the classification

To assess the performance of the model on the test datasets, in this paper we are used the following accuracy (ACC) measure to estimate the classifier accuracy of the proposed algorithm. The accuracy is defined as follows:

$$ACC = [(T_1 + T_2 + T_3) / N] * 100\% \quad (10)$$

Where  $T_1$ ,  $T_2$ , and  $T_3$  are the number of correctly classified normal, benign and malign cases; respectively.  $N$  is the total number of test samples. Table (2) shows the accuracy results of the classification algorithm.

Folder	Number of rules	Accuracy
1	22	97.2
2	18	82.8
3	22	86.3
4	22	82.8
5	40	76.7
6	34	79
7	32	98.3
8	25	89.3
9	16	73.9
10	17	88.3
Avg(%)	24.8%	85.46%

Table 2: Classification accuracy over the 10 folder

## 8. Conclusion

Mammography is one of the best methods in breast cancer analysis, but in some cases, radiologists can not analysis tumors despite their experiences. Such computer-aided methods like those presented in this paper could assist medical staff and improve the accuracy of detection. This paper presents an efficient classification based on texture features to classify from medical databases in the context of rough set theory. Five features generated form the co-occurrence matrix are extracted and represented in attribute vector, and then the decision rules within the data are extracted. Therefore, the classifier model was built and the quadratic distance similarly is used for matching process. The experimental results show that the algorithm performs well reaching over 85% in accuracy.



## References

1. Aboul Ella H., "Rough Sets Analysis for Reduct Generation of Breast Cancer Data" International Journal of Artificial Intelligent Tools, IJAIT, Scientific World Publisher, 2003 (to appear).
2. Aboul Ella H. and Jafar M. H. Ali, "An Efficient Classification and Image Retrieval Algorithm based on Rough Set Theory" 5th International Conference On Enterprise Information Systems École Supérieure d' Électronique de l' Ouest - Angers - France - 23-26 April, pp. 141-147,2003.
3. Aboul Ella H. Taha E.A. and Tarek A. H., "Digital Mammography Image Analysis System Based on Mathematical Morphology" IEEE computer society 7th International Conference on Intelligent Engineering Systems INES 2003, March 4- 6 Assiut-Luxor, Egypt, 2003.
4. Ahuja N. and Rosefeld A., "A Note on the Use of Second-order Gray-level Statistics for Threshold Selection" IEEE Trans. Systems, Man, and Cybernatics, SMC-8912, pp. 895-898, 1978.
5. Cancer Net Home Page of the National Cancer Institute <http://biomed.nus.sg>, 2002.
6. Dougherty E. R., "Mathematical Morphology in Image Processing" Marcel Dekker, New York, 1993.
7. Grzymala-Busse J., Pawlak Z., Slowinski R. and Ziarko W., "Rough Sets", Communications of the ACM, vol.38, no. 11, 1999.
8. Hand D.J., Mannila H. and Smyth P., "Principles of Data Mining" MIT Press, 2000.
9. Haralick R.M., "Statistical and Structural Approaches to Texture" Proceeding of the IEEE, vol.67, no.5, May 1979.
10. <http://www.wiau.man.ac.uk/services/MIAS/MIASmini.htm>, "The Mammographic Image Analysis Society: Mini Mammography Database" 2003.
11. Kok S.L., "Computer Interpretation of X-ray Mammography" Department of Engineering Science, University of Oxford, web site <http://www.eng.ox.ac.uk>, 2003.
12. Lavrac N., "Machine Learning for Data Mining in Medicine" In AIMDM'99, pp.47-64, 1999.
13. Liu S. Babbs C.F. and Delp E.J., "Normal Mammogram Analysis and Recognition" Proceedings of the IEEE International Conference on Image Processing, October 4-7, Chicago, Illinois,1998.
14. Mari P., Bogdan C. Moncef G. and Ari V., "Rock Texture Retrieval Using Gray Level Co-Occurrence Matrix". NORSIG-2002, 5th NORDIC Signal Processing Symposium, Oct. 4-7, Norway, 2002.
15. Martin J. Yaffe "Digital Mammography" The fifth International Wrkshops on Digital Mammography, IWDM'2000, June 11-14, Toronto, Canada, 2000.
16. Nico K. Martin T. Jan H. and Leon V. E., "Digital Mammography: Computational Imaging and Vision" Kluwer academic publication, 1998.
17. Pawlak Z., "Rough sets" Int. J. Computer and Information Science, vol. 11, pp. 341-356, 1982.
18. Slowinski R., "Rough Set Approach to Decision Analysis" AI Expert, pp. 19-25, March 1995.