IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

# Effect of Image Mining Technique to Detect Breast Cancer from Mammograms synthesizing Rough Set Theory and Association Rule Approach

**Aswini kumar mohanty**

Capital engineering college, khurda

Abstract: - Image mining relates and an exhibition reference to a data mining technique where images are used as data. Image mining technique can extract knowledge and exciting patterns that are not stored in the database by analyzing the images using various tools. Image data mining is one of the core functions of this current scenario as image data plays a vital role in every aspect of the system. Breast cancer is a ubiquitous common cancer in women, as well as in several men. Mammography is the first imaging method to detect breast cancer, although it is highly dependent on the radiologist's diagnosis. Computer aided diagnosis has advanced a lot to diagnose cancer correctly. In this paper, the author uses a novel and time-saving approach to breast cancer diagnosis using rough set theory and association rule mining. The accuracy level of the system is quite acceptable and can help the doctor easily detect breast cancer without double reading with guaranteed results. The accuracy level reaches above 98% which can be acceptable and the time complexity is also too less. The main objective of this paper is to apply image mining on mammography images to classify and detect cancerous tissue by discovering insights from the hidden pattern of cancerous tissue. Data mining extended to image fields are techniques generally more suitable for larger databases. The features need to be calculated and their correct subset should significantly increase the classification accuracy. Association rule algorithms generally adopt an iterative method to discover frequent itemsets, which requires very extensive calculations and a complicated transaction process. For this reason, a modified algorithm of association rules is proposed in this paper. Experimental results show that this method can quickly discover frequent item sets and efficiently mine potential association rules. Texture features are the most predictive features, which include intensity histogram features and GLCM features that are extracted from mammography images containing knowledge discovery attributes. In this work, a new feature subset selection approach using Rough Set Improved Harmony Search Quick Reduct is proposed, which reduces approximately 50-60% of the features, and the proposed association rule is used for classification. Most interestingly, the Rough Set Improved Harmony Search Quick Reduct (RSIHSQR) algorithm provides the best optimal features. The experiments were taken from a dataset of 322 MIAS images of various types with the aim of improving accuracy by generating a minimum numbers of rule to cover multiple patterns.

**Keywords:** Mammogram, Gray Level Co-occurrence Matrix feature, Histogram Intensity, Feature selection based on Rough Set Improved Harmony Search Quick Reduct (RSIHSQR), Association rule mining.

## 1. INTRODUCTION

Breast cancer is one of the most common cancers, a widespread cause of death among women, especially in developed countries and now also in developing countries. There is no primary prevention because the cause is still a mystery. Thus, early detection of cancer stage can facilitate treatment that could lead to better survival rates. Mammography is currently the primary imaging method to detect breast cancer screening. However, 10–30% of breast cancers are missed by mammography [1]. Image Mining is the best way to extract hidden patterns of information and knowledge from a large database of images. It 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 [2,3,4,5,6,7]. The generalized image

mining technique for knowledge discovery is illustrated in fig.1. Similarly the methodology for classifying breast tumor using image mining technique is depicted in fig 2.

Image data mining of medical images is used to collect effective models, relationships, rules, abnormalities, and patterns from large volumes of data. This procedure can speed up the diagnosis and decision-making process. Various data mining methods such as wavelet [8,9], statistical methods and most of them used features extracted using image processing techniques [10] have been used to detect and classify anomalies in mammographic images. Some other methods are based on fuzzy theory. [11,12] and neural networks [9]. More specifically, in this paper, we propose a novel associative classifier that selects "strong" class association rules based on the overall coverage of the training set. The advantage of the proposed classifier is that it generates significantly smaller rules for larger datasets compared to traditional classifiers while maintaining classification accuracy. The result shows that the proposed rule-based approach achieves a classification accuracy of over 97% and also demonstrates the use and effectiveness of association rule mining in image classification [13-16].

Classification and prediction are two important data mining tasks that can be used to extract hidden regularities from given datasets to build accurate models (classifiers). Classification approaches focus on creating classifiers (models) for predicting the class designation of a future data object. Such analysis can help us understand the data better in a comprehensive way. Classification methods have been widely used in many real-world applications, such as customer relationship management [17], medical diagnostics [18], and industrial design [19]. The classification process usually involves two phases: a training phase and a testing phase. In the training phase, the features of the typical features of the image are isolated and based on this training class is created. In the subsequent testing phase, these partitions of the feature space are used to classify the image. We used the supervised genetic association rule method by extracting low-level image features for classification. The advantage of this method is efficient character extraction, selection and effective classification. The remainder of the paper is organized as follows. Section 2 presents the breast tumor morphology (ROI) followed by preprocessing. Section 3 presents the histogram equalization and feature extraction stage. Section 4 discusses the methodology description and Section 5 presents the feature selection and classification method, followed by Section 6. Section 7 discusses the results and Section 8 concludes.

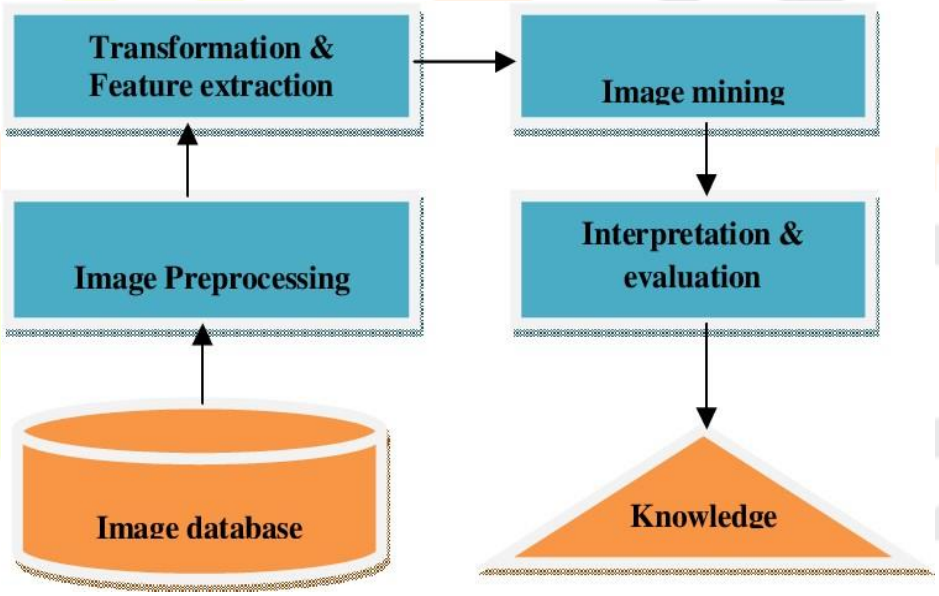


Fig.1 generalized image mining technique for knowledge discovery

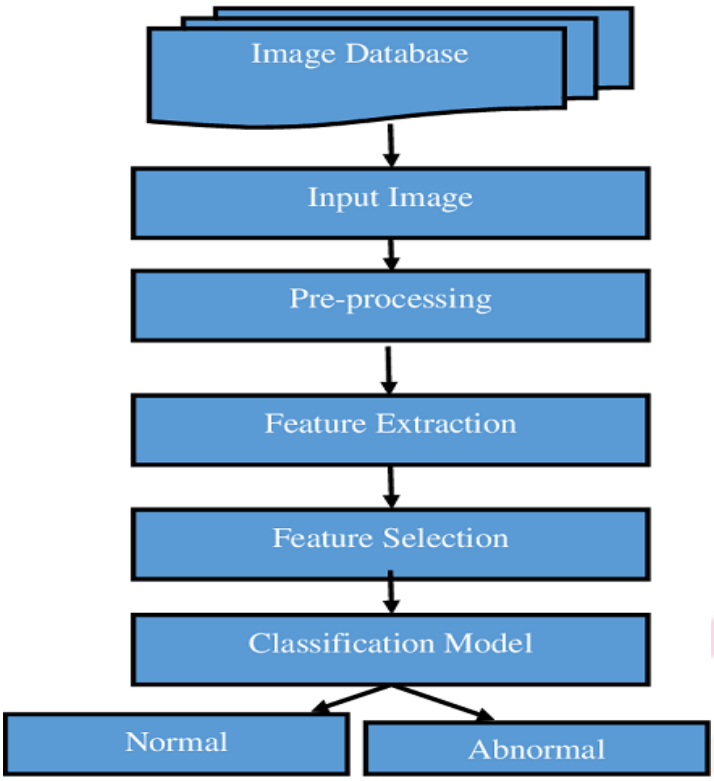


Fig.2 Image mining framework for classification of Breast cancer

## 2. PRE-PROCESSING

The mammography image for this study was procured from the Mammography Image Analysis Society (MIAS), a British breast cancer research group [20]. Since mammograms are difficult to interpret, preprocessing is necessary to improve image quality and make the feature extraction phase easier and more reliable. A cluster of calcifications/tumor is surrounded by breast tissue that masks the calcifications, preventing accurate detection and is shown in Figures 2.1 .A preprocessing; a noise reduction step is usually used to improve image contrast and calcification. In this work [21], an effective low-pass filter was applied to the image, which retained calcifications while suppressing unimportant image features.

Figures 4 shows representative output image of the filter for a image cluster in figure 3. By comparing the two images, we observe background mammography structures are removed while calcifications are preserved. This simplifies the further tumor detection step

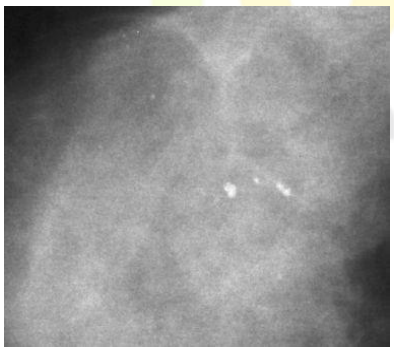

Fig 3 ROI of a Benign

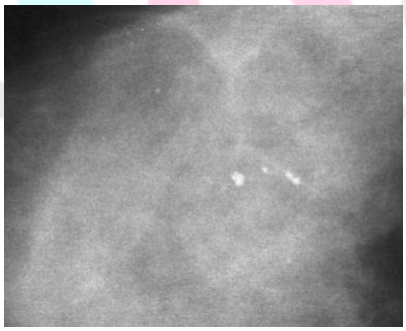

Fig. 4.ROI after Pre-processing Operation

### 2.1. Histogram equalization

Histogram equalization is a method in image processing of contrast adjustment using the image histogram [22]. Thanks to this adjustment, the intensities can be better distributed on the histogram. This allows areas with lower local contrast to gain better contrast. Histogram smoothing achieves this by effectively distributing the most frequent intensity values. The method is useful for images with a background and foreground that are light or both dark. In particular, this method can lead to better visualization of bone structure on X-ray images and better detail in photos that are overexposed or underexposed. In mammography images, histogram equalization is used to adjust contrast to make image abnormalities more visible.

## 3. FEATURE EXTRACTION

Properties, characteristics of objects of interest, if carefully selected, represent the maximum relevant information that the image can offer for a complete characterization of the lesion [23,24]. A feature extraction methodology analyzes objects and images to extract the most significant features that are representative of different classes of objects. Elements are used as inputs to classifiers that assign them to the class they represent. In this, the labor intensity histogram and gray level co-occurrence matrix (GLCM) functions are extracted.

### 3.1. Intensity Histogram Features

Histogram intensity analysis was extensively investigated in the early stages of the development of this algorithm [25]. Previous studies have provided intensity histogram properties such as mean, variance, entropy, etc. These are summarized in the table. 1 Mean values characterize individual calcifications; Standard deviations (SD) characterize the cluster. Table 2 summarizes the values for these functions.

Table.1 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 2 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 takes into account the spatial relationship of pixels, namely the gray-level co-occurrence matrix (GLCM), also known as the gray-level spatial dependence matrix [26,27]. By default, the spatial relationship is defined as the pixel of interest and the pixel to its right (horizontally adjacent), but you can specify other spatial relationships between the two pixels. Each element (I, J) in the resulting GLCM is simply the sum of the number of times a pixel with value I occurred in a 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 correlation1, information measure of correlation2, Inverse difference normalized. Information difference normalized. The value obtained for the above features from our work for a typical image is given in the following table 3

Table 3 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 correlation1 | 1.2145 |
| 18 | Information measure of correlation2 | -0.0322 |
| 19 | Inverse difference normalized | 0.2863 |
| 20 | Information difference normalized | 0.9107 |

## 4. FEATURE SELECTION

Feature selection helps reduce feature space, which improves prediction accuracy and minimizes computation time [28]. This is calculated and computed by removing irrelevant, redundant and noisy features, i.e. selecting a subset of features that can achieve the best performance in terms of accuracy and computation time. Performs dimension reduction. A number of search procedures have been proposed. Popular feature selection algorithms used are sequential forward selection, sequential backward selection, genetic algorithm, and particle swarm optimization, branch and bound optimization. In this work, the author uses a new feature selection method based on rough set theory called Rough Set Improved Harmony Search Quick Reduct (RSIHSQR), which works on the discretization value and at the same time improves the learning process of prediction methods[29,30]. In this work, the RSIHSQR feature selection algorithm is implemented to reduce the redundant and intangible features of the available datasets and reveal the best features. This technique is implemented for a large data set and has shown plausible results in protein analysis [31,32]. The spirit of the rapid rough set reduction method in the proposed algorithms is inherited in the improved HS algorithm to converge to the best solution, and the solution is achieved by the musical improvisation technique [33,34], where the algorithm is briefly described in [35,36]. The results of the algorithm that selects the best features are used as input to a classifier for analyzing and classifying mammograms.

### 4.1. Discretization of 1st order and 2nd order GLCM Features

The continuous value dataset with real-valued attributes is discretized for an effective classification using association rule as association rule mining works on binary value [37,38]. Discretization is the process of transferring continuous values to binary values representing one and zero, and rough set theory works on discrete variables to deal with the dilemma of attributes presented in the real dataset [39]. The mammogram GLCM feature dataset experimented within this study has real-valued attributes. Hence, it is discretized by applying the Max-Min discretization method [40,41] and then it was applied for experiment to select relevant features by feature selection algorithms in the proposed framework.

**Algorithm 1.** ***RSIHSQR FS algorithm***

***Algorithm: RSIHSQR(C,De)***

***Input: C, the conditional attributes;***

***De, the class attribute;***

***Output: Optimal feature subset***

*Step I: The fitness function, f(X) to be defined.*

*Initialize the parameters HMS = 30*

*HMCR = 0.9 // HM constraint*

*MaxIt = 100 // iteration count*

*PVB//feasible value limit*

$PAR_{min}, PAR_{max}, bw_{min}, bw_{max}$, *// Pitch Adjusting Rate & bandwidth* $\in$ *(0 to 1)*

*fit = 0;*

$X_{old} = X_1$; *bestfit* $= X_1$; *bestreduct = {};*

*Step II: The Harmony Memory to be set as,* $HM = (X_1, X_2, \ldots, X_{HMS})$

*For i = 1:HMS*

$\forall$: $X_i$ *// *$X_i$ *is the ith harmony of HM*

$R \leftarrow X_i$ *(1's of* $X_i$*)*

$\forall x \in (C - R)$

$$\gamma_{R\cup\{x\}}(De) = \frac{|POS_{R\cup\{x\}}(De)|}{|U|}$$

$f(X_i) = \gamma_{R\cup\{x\}}(De)$ *for all* $X \subset R$, $\gamma_X(De) \neq \gamma_C(De)$

*if* $f(X_i) >$ *fit*

*fit* $\leftarrow f(X_i)$

$X_{old} \leftarrow X_i$

*End if*

*End for*

*Step III: The new HM to be improvised.*

*While itr* $\leq$ *MaxIt |* *fit == 1*

*for j = 1:NVAR*

$\forall: X_{old}(j)$

*Update PAR();*

*Update bw();*

*if random ( )* $\leq$ *HMCR* *//random numberbetween 0 and 1*

$X_{new} \leftarrow X_{old}$;

*if random ( )* $\leq$ *PAR*

$X_{new}(j) = X_{new}(j) \pm random() * bw$

*end if*

*else*

*// select Xnew*

$X_{new}(j) = PVB_{lower} + random( ) * (PVB_{upper} - PVB_{lower})$

*end if*

*end for*

*Step IV: The new HM to be updated*

*Calculate fitness for* $X_{new}$*—(Step II)*

*if* $f(X_{new}) \geq f(X_{old})$

*// Substitute the older harmony with new harmony, if it is best*

$X_{old} \leftarrow X_{new}$;

*if* $f(X_{new}) >$ *fit*

*fit* $\leftarrow f(X_{new})$;

*bestfit* $\leftarrow X_{new}$;

*End if*

*Exit*

*end if*

*end while*

*bestreduct* $\leftarrow$ *selected attributes of bestfit*

Table 4. Feature selected by RSIHSQR feature selection 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 |
| 10 | Entropy |
| 11 | Difference variance |
| 12 | Difference entropy |
| 13 | Max probability |
| 14 | Sum of variance |

## 5. CLASSIFICATION

Creating accurate classifiers is one of the fundamental tasks of data mining and machine learning research. Given a set of training instances with known class labels, classifiers aim to predict target classes for a set of test instances for which the class labels are unknown. First, a classification model is built from the training data and then used to classify the unseen cases. There are various methods for building classifiers, such as decision trees [42], naive-Bayesian methods [43], statistical approaches [44], support vector machines [45], etc.

The association classification (AC) technique is a predictive approach that has been widely investigated in the last two decades. Many researchers have tried to use AC in real-world applications such as: text classification, medical diagnosis, fraud detection, and website phishing. However, there are several concerns about using this technique and they are as follows: first, it generates too many rules that consume a lot of time and memory compared to classical data mining techniques. Second, the support and confidence thresholds are estimated by the user and therefore the evaluation process is always influenced by these values. To overcome these problems, we proposed a modified technique to enhance the performance of the Apriori algorithm used in association-based classification (CBA). A technique we discovered that works well involves using the harmonic mean to generate more reliable rules and to improve the evaluation process of the generated rules.

### 5.1 Finding the class association rules

The step of finding frequent sets of items in associative classification is the most important and computationally expensive step [46,47]. Since association rule discovery, several different approaches have been taken to discover frequent rule entries from a dataset. For example, some AC methods [48,49] use the "APRIORI" candidate generation method. Other AC methods [50,51,52] mainly use the FP-growth approach [53] to generate association rules. The CPAR algorithm uses the greedy strategy developed in FOIL [54]. This portion of the section urges on discovering class association rules from frequent itemsets.

Generating association rules is usually divided into two main steps:

1. In the first step, a minimum support threshold is used to find all frequent itemsets from the training dataset.

2. A minimum confidence constraint is used to generate strong class association rules from frequent itemsets.

In the first step, we apply the "APRIORI" algorithm to find frequent itemsets. "APRIORI" uses the "downward-closure" property to speed up the search process by reducing the number of candidate sets of items at any level. The main feature of "top-down closure" is that all subsets of frequent items must be frequent. Uncommon itemsets found at any level are removed, as it is not possible to "create" a common itemset from an uncommon one. "APRIORI" uses this process to prune infrequent itemsets before calculating their support at any level. This reduces the time complexity of calculating the support for all combinations of items in the data set.

Once all frequent itemsets from the learning dataset are found in the first step, a straightforward approach is to generate strong class association rules that satisfy both the minimum support as well as confidence of minimal constraints from these frequent itemsets in the second step. This can be done using the following equation for confidence:

$$confidence(X \rightarrow C) = \frac{support\_count(X \cup C)}{support\_count(X)} \quad (1)$$

Equation (1) is expressed as the support count of itemsets, where X is the antecedent (the itemsets, i.e. the left side of the rule), C is the consequence (the class label, which is the right side of the rule), support_count (X ∪ C) is the number of transactions containing itemsets X ∪ C and support_count (X) is the number of transactions containing items A. Based on this equation, CAR can be generated as follows:

(a) For each frequent itemsets *L* and class label *C*, generate all nonempty subsets of *L*.

(b) For every nonempty subset S of L, output the strong rule R in the form of "S→C" if, $\frac{support\_count(R)}{support\_count(S)} \geq$ min_con f , where min_conf is the minimum confidence threshold.

## 6. Proposed Associative Classification Approach—J&B

Our proposed approach—J&B can be defined in three steps:
1. A complete set of strong class association rules is generated from the training dataset discussed in Section 4.

2. Then we select strong class association rules that contribute significantly to increase the coverage of the classifier until we meet the stopping criterion to produce an accurate, compact and understandable classifier.

3. In the last step, we find the overall coverage and accuracy of the intended classifier to show how the overall coverage of the classifier affects its classification accuracy.

### 6.1. Proposed Intended Classifier

Once we have generated strong class association rules in Section 4, we will design our classifier using them. The method proposed by us is described in Algorithm 1.

Lines 1–2 find all frequent itemsets in the training dataset using the "APRIORI" algorithm described in Section 5. The third line generates strong class association rules that satisfy the minimum support and confidence constraints from the frequent itemsets. The class association rules (generated on line 3) are ordered by confidence and support descending order in the fourth line according to the following criteria: Given two rules R1 and R2, R1 is said to have a higher rank than R2, denoted as R1 > R2,

• If and only if, con f (R1) > con f (R2); or

• If con f (R1) = con f (R2), but, supp(R1) > supp(R2); or

• If con f (R1) = con f (R2) and supp(R1) = supp(R2), R1 has fewer attribute values on the left than R2;

• If all parameters of the rules are the same, we can choose any of them.

The initial values of the traindata_classified field are filled with a false value in line 5. This field is later used to calculate the total coverage of the training data. Basically, the classifier is built on lines 6-17. We scan the training dataset for each rule (lines 6–7): if the rule classifies the new unclassified example as 5 (that is, the body of the rule matches the body of the new example), we increase the contribution (that is, the rule to increase the total coverage) in lines 8–11 . If the contribution is greater than 0, meaning the rule has classified the new example(s), we increase the total coverage by the

contribution and add this rule to our final classifier on lines 12–14. If a rule can classify the training examples but cannot contribute to improving the overall coverage, then we do not evaluate that rule as a potential rule for our classifier. Lines 15–16 stop the procedure if the classifier reaches an intended total coverage that is greater than the user-defined intended_coverage threshold, and line 17 returns the proposed classifier.

**Algorithm 1:** Building a compact and accurate classifier—J&B

**Input:** A training dataset D, minimum support and confidence constraints, intended coverage
**Output:** A subset of strong class association rules for classification

```
1:  D = taining_dataset;
2:  F = frequent_itemsets(D);
3:  Rule = genCARs(F);
4:  Rule = sort(R, minconf, minsup);
5:  Fill(classified_traindata, false):
6:  for i: = 1 to Rule.length do begin
7:    for j: = 1 to D.length do begin
8:      if classified_traindate[j] = false then
9:        if Rule[i].premise classifies D[j].premise then
10:           classified_traindate[j] = true;
11:           contribution = contribution + 1;
12:   if contribution > 0 then
13:     overall_coverage = overall_coverage + contribution;
14:     Classifier = Classifier.add(rule[i]);
15:   if (overall_coverage/D.length) ≥ intended_coverage then
16:     break;
17: return Classifier
```

The classification process of our method is shown in Algorithm 2.

**Algorithm 2:** Classification process of J&B

**Input:** A Classifier and test dataset
**Output:** Predicted class

```
1:  for each rule y ∈ Classifier do begin
2:    if y classify test_example then
3:      class_count[y.class]++;
4:    if max(class_count) = 0 then
5:      predicted_class = majority_classifier;
6:    else
7:      predicted_class = max_index(class_count);
8:  return predicted_class
```

Algorithm 2 depicts the procedure of using the J&B classifier to find the class label of a (new) test example. For each rule in the J&B classifier (line 1): if the rule can classify the example correctly, then we increase the corresponding class count by one and store it (lines 2–3). If none of the rules can classify the new example correctly, then the algorithm returns the majority class value (lines 4–5);otherwise, it returns the majority class value of correctly classified rules (lines 6–7).

### 6.2. *Overall Coverage and Accuracy*

After building our classifier (as in Section 5.1), it is easy to calculate the overall coverage and accuracy of the classifier (defined in Algorithm 3). To calculate the total coverage, we count the examples that the classifier covers and divide by the total number of examples in the dataset. For accuracy, we count all examples that the classifier classifies and divide by the total number of examples in the dataset.

The first line finds the length of the dataset. On the second and third lines, we fill all the initial values of fieldclassified_test_data (to calculate the overall coverage) and classified_test_data_withclass (to calculate the overall accuracy) as false. Lines 4–11 generally toggle the state of the test example, with no class label for calculating total

coverage in lines 6–8 and with a class label for calculating precision in lines 9–11. The number of correctly classified examples by the classifier without class labels and with class labels is counted in lines 12–16 to calculate the overall classifier coverage, and the overall accuracy is calculated in lines 17–18. The last line returns the results obtained.

**Algorithm 3:** Overall coverage and accuracy of the classifier

```
Input: Dataset D and classifier C
Output: Overall coverage and accuracy of the classifier
1:  n = D.length();
2:  Fill (classified_testdata, false);
3:  Fill (classified_testdata_withclass, false);
4:  for i: = 1 to C.length do
5:    for j: = 1 to n do
6:      if C[i].premise classifies D[j].premise then
7:        if classified_testdata[j] = false then
8:          classified_testdata[j] = true;
9:      if C[i] classifies D[j] then
10:       if classified_testdata_withclass[j] = false then
11:         classified_testdata_withclass[j] = true;
12: for i: = 1 to n do
13:   if classified_testdata[i] then
14:     testcoverage++;
15:   if classified_testdata_withclass[i] then
16:     accuracy++;
17: Overallcoverage_testdata = testcoverage/n;
18: Overallaccuracy_testdata = accuracy/n;
19: return Overallaccuracy_testdata, Overallcoverage_testdata
```

## 7. Experimental Evaluation

We tested our model for J&B evaluation, which was run with default parameters of minimum support = 1% and minimum confidence = 60%. However, the minimum support has been reduced to 0.5% or even 0.1% to ensure enough CAR to generate for each class value. We used their WEKA workbench implementation [55] with default parameters. We used 90% as the required coverage threshold (training data set) which is the stopping point for selecting rules for our classifier.

Average accuracy is 98.44%. We used precision, recall, and F-measures as evaluation metrics for mammogram classification. Accuracy is the fraction of true positive predictions divided by the total number of true positive predictions in the set. Recall is the total number of prediction outcomes divided by the total number of true positives in the set. F-Measure is the provision and a way to combine precision and recall into a single measure that captures both properties. F-Measure = (2 * Precision * Recall) / (Precision + Recall). The result of testing using the selected features is used for the classification shown in Table 5. To classify the samples, we used the freely available Machine Learning package, WEKA [55] to train our dataset by the J&B method. Of the 322 images in the dataset, 210 were used for training and the remaining 102 for testing purposes.

Table 5. Result of proposed method

| | |
|---|---|
| Normal | 99% |
| Malignant | 94.11% |
| Benign | 100% |

The confusion matrix has been obtained from the testing part .In this case out of 51 actual malignant images 03 images was classified as normal. In case of benign all images are correctly classified but in case of normal 207 images are correctly classified on average of 98.44% accuracy. The confusion matrix is given in Table 8. The performance of the propped classifier is shown in figure 6.

Table 6. Confusion matrix

| Actual | Predicted class | | |
|---|---|---|---|
| | Benign | Malignant | Normal |
| Benign | 62 | 0 | 0 |
| Malignant | 51 | 48 | 03 |
| Normal | 209 | 2 | 207 |

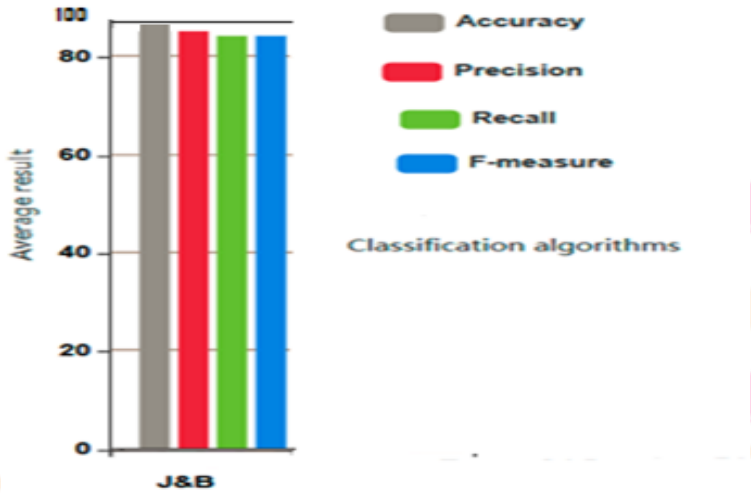


Fig.5 J&B classifier's Accuracy, Precision, Recall and F-measure

## 8. CONCLUSION

Image mining is one of the best methods to draw the relationship between different attributes and find relevant patterns in an image from which knowledge can be discovered. Breast cancer detection and diagnosis has been studied for more than two decades and has still not been a reliable diagnostic tool for the radiologist. Researchers are constantly striving to improve the accuracy of diagnosis by reducing false positives and false negatives. Although the FDA has approved computer-aided detection (CAD) for breast cancer, both the oncologist and the radiologist still have the dilemma of whether to use CAD because of the highest false-positive rate. However, we proposed a different kind of technique by integrating rough set theory and association rules that use a discretized value for feature selection and classification. The methodology adopted by the author is undoubtedly very flexible and supports fast execution on a large database, and extracting the most meaningful features significantly contributes not only to increasing the classification accuracy, but also reduces the time consumption and also achieves the standard accuracy. With the help of a radiologist to crop a mammographic image that raises doubts about both microcalcification and cancerous tissue, these gray images are taken into account without using a segmentation technique. Both statistical and GLCM features are more important than other image features for pattern recognition.

J&B produces classifiers that have, on average, far fewer rules than those produced by "classical" rule learning methods, and a slightly smaller number of rules compared to the associative classification models included in the comparison. J&B generated a reasonably smaller number of rules for larger datasets compared to all other classification methods. In this research work, one of the new approaches is to use the intended coverage constraint, which becomes the rule selection criterion for our classifier. Here we tried to show that it is possible to achieve reasonably high accuracy without covering the entire dataset (especially when the dataset contains some missing or noisy examples). This criterion becomes an advantage of our method, which creates significantly smaller classification rules on larger datasets. J&B created an accurate and compact classifier on a dataset that has a particularly large number of attributes and records, with balanced class distribution datasets. The proposed approach is applied to the construction of an association rule for classifying

images into three classes: normal, benign, and malignant. The result shows that it outperforms the existing one. In the future, an efficient algorithm can be used to select relevant features and generate rules to improve accuracy.

## 9. REFERENCES


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

[2] Osmar R. Zaïane,M-L. Antonie, A. Coman "Mammography Classification by Association Rule based Classifier," MDM/KDD2002 International Workshop on Multimedia Data Mining ACM SIGKDD, pp.62- 69,2002,

[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, pp7459 – 7462, 2005.

[4] C.Chen and G.Lee, "Image segmentation using multitiresolution wavelet analysis and ExpectationMaximum(EM) algorithm for mammography" , International Journal of Imaging System and Technology,8(5): pp491-504,1997.

[5] Ji Zhang, Wynne Hsu, Mong Li Lee, "Image mining: Issues, Frame works and Techniques', National University of Singapore.

[6] Osmar R. Zaiane, Jiawei Han, Ze-Nian Li, Hean Hou, "Mining Multimedia data", Simon Fraser University, Canada.

[7] Chabane Djeraba, "Relationship Extraction from Large Image Databases", University of Nautes, France.

[8] "Knowledge Extraction and Learning Object Mining", University of Waterloo.

[9] Scott E. Umbaugh, "Computer Vision and Image Processing: A Practical Approach Using CVIP tools".

[10] 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,pp. 36 – 41, 2005.

[11] Jensen, Qiang Shen, "Semantics Preserving Dimensionality Reduction: Rough and Fuzzy-Rough Based Approaches", IEEE Transactions on Knowledge and Data Engineering, pp. 1457-1471, 2004.

[12] I.Christiyanni et al ., "Fast detection of masses in computer aided mammography", IEEE Signal processing Magazine, pp:54-64,2000.

[13] K. Thangavel , A. Kaja Mohideen "Classification of Microcalcifications Using Multi-Dimensional Genetic Association Rule Miner" International Journal of Recent Trends in Engineering, Vol 2, No. 2, pp. 233 – 235, 2009

[14] R.Agrawal, T. Imielinski, and A.Swami. Mining association rules between sets of items in large databases. In the Proc. of the ACM SIGMOD Int'l Conf. on Management of Data (ACM SIGMOD '93), Washington, USA, May 1993.

[15] Alex A. Freitas, "A Survey of Evolutionary Algorithms for Data Mining and Knowledge Discovery" Postgraduate Program in Computer Science, Pontificia Universidade Catolica do Parana Rua Imaculada Conceicao, 1155. Curitiba - PR. 80215-901. Brazil.

[16] Etta D. Pisano, Elodia B. Cole Bradley, M. Hemminger, Martin J. Yaffe, Stephen R. Aylward, Andrew D. A. Maidment, R. Eugene Johnston, Mark B. Williams,Loren T. Niklason, Emily F. Conant, Laurie L. Fajardo,Daniel B. Kopans, Marylee E. Brown • Stephen M. Pizer "Image Processing Algorithms for Digital Mammography: A Pictorial Essay" journal of Radio Graphics Volume 20,Number 5,sept.2000

[17] Eric, N.; Xiu, L.; Chau, D. Application of data mining techniques in customer relationship management.A literature review and classification. Expert Syst. Appl. 2009, 36, 2592–2602.

[18] Jyoti, S.; Ujma, A.; Dipesh, S.; Sunita, S. Predictive data mining for medical diagnosis: An overview of heart disease prediction. Int. J. Comput. Appl. 2011, 17, 43–48.

[19] Yoon, Y.; Lee, G. Two scalable algorithms for associative text classification. Inf. Process Manag. 2013, 49, 484–496.

[20] Etta D. Pisano, Elodia B. Cole Bradley, M. Hemminger, Martin J. Yaffe, Stephen R. Aylward, Andrew D. A. Maidment, R. Eugene Johnston, Mark B. Williams,Loren T. Niklason, Emily F. Conant, Laurie L. Fajardo,Daniel B. Kopans, Marylee E. Brown • Stephen M. Pizer "Image Processing Algorithms for Digital Mammography: A Pictorial Essay" journal of Radio Graphics Volume 20,Number 5,sept.2000

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

[22]. 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; pp 194–202, 1995.

[23]. Yvan Saeys, Thomas Abeel, Yves Van de Peer Towards robust feature selection techniques, www.bioinformatics.psb.ugent

[24] Gianluca Bontempi, Benjamin Haibe-Kains Feature selection methods for mining bioinformatics data, http://www.ulb.ac.be/di/mlg

[25] Li Liu, Jian Wang and Kai He Breast density classification using histogram moments of multiple resolution mammograms Biomedical Engineering and Informatics (BMEI), 3rd International Conference, IEEE explore pp.146–149, DOI: November 2010, 10.1109/ BMEI.2010 .5639662,

[26] Li Ke,Nannan Mu,Yan Kang Mass computer-aided diagnosis method in mammogram based on texture features, Biomedical Engineering and Informatics (BMEI), 3rd International Conference, IEEE Explore, pp.146 – 149, November 2010, DOI: 10.1109/ BMEI.2010.5639662,

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

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

[29] Cao, Y.; Liu, S.; Zhang, L.; Qin, J.;Wang, J.; Tang, K. Prediction of protein structural class with Rough Sets.

BMC Bioinform. 2006, 7,20.
[30] Anaraki, J.R.; Eftekhari, M. Rough set based feature selection: A Review Fifth conference on information and knowledge technology (IKT). IEEE 2013, 2013, 301–306.
[31] ,Bagyamathi, M.; Inbarani, H.H. Prediction of Protein Structural Classes by Pseudo Amino Acid Composition Using Improved Harmony Search Relative Reduct Feature Selection and Rough Set Classification Algorithms. Int. J. Invent. Comput. Sci. Eng. 2017, 4, 55–65.
[32] Bagyamathi, M.; Inbarani, H.H. Prediction of Protein Structural Classes using Rough Set based Feature Selection and Classification Framework. J. Recent Res. Eng. Technol. 2017, 4, 1–9.
[33] Poursalehi, N.; Zolfaghari, A.; Minuchehr, A. Differential harmony search algorithm to optimize PWRs loading pattern. Nucl. Eng. Des. 2013, 257, 161–174.
[34] Geem, Z.W. Particle-Swarm Harmony Search for Water Network Design. Eng. Optim. 2009, 41, 297–311.
[35] Bagyamathi, M.; Inbarani, H.H. A Novel Hybridized Rough Set and Improved Harmony Search Based Feature Selection for Protein Sequence Classification. Big Data in Complex Systems: Challenges and Opportunities, Studies in Big Data; Springer: Berlin/Heidelberg, Germany, 2015; Volume 9, pp. 173–204.
[36] Bagyamathi, M.; Inbarani, H.H. Feature Selection using Improved Harmony Search Hybridized with Relative Reduct for Medical Data Classification. Int. J. Appl. Eng. Res. (IJAER) 2015, 10, 19476–19480.
[37] Beniwal, S.; Arora, J. Classification and feature selection techniques in data Mining. Int. J. Eng. Research Technol. 2012, 1, 2278–2284.
[38] Kotsiantis, S.; Kanellopoulos, D. Discretization Techniques: A recent survey. GESTS Int. Trans. Comput. Sci.Eng. 2006, 32, 47–58.
[39.] Ali, R.; Siddiqi, M.H.; Lee, S. Rough set-based approaches for discretization: A compact review. Artif Intell. Rev. 2015, 44, 235–263.
[40] Tsoukalas, A.; Parpas, P.; Rustem, B. A smoothing algorithm for finite min–max–min problems. Optim. Lett. 2008, 3, 49–62.
[41] Sathishkumar, E.N.; Thangavel, K.; Nishama, A. Comparative analysis of discretization methods for gene selection of breast cancer gene expression data. In Computational Intelligence, Cyber Security and Computational Models; Springer: Coimbatore, India,2014; pp. 373–378.
[42] J. R. Quinlan. C4.5: Programs for Machine Learning. Morgan Kaufmann, 1993.
[43] R. Duda and P. Hart. Pattern Classification and Scene Analysis. JohnWiley & Sons, 1973.
[44] T. S. Lim, W. Y. Loh, and Y. S. Shih. A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. Machine Learning, 39, 2000.
[45] Cristianini, N., Shawe-Taylor. An Introduction to Support Vector Machines, Cam- bridge University Press, (2000).
[46] Yin, X.; Han, J. CPAR: Classification based on Predictive Association Rules. In Proceedings of the SIAM International Conference on Data Mining, San Francisco, CA, USA, 1–3 May 2003; Barbara, D., Kamath C.,Eds.; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA; pp. 331–335.
[47] Hu, L.Y.; Hu, Y.H.; Tsai, C.F.; Wang, J.S.; Huang, M.W. Building an associative classifier with multiple minimum supports. SpringerPlus 2016, 5, 528.
[48]. Chen, G.; Liu, H.; Yu, L.; Wei, Q.; Zhang, X. A new approach to classification based on association rule mining. Decis. Support Syst. 2006, 42, 674–689.
[49] Meretakis, D.; Wüthrich, B. Classification as Mining and Use of Labeled Itemsets. In Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA,15–18 August 1999; Fayyad, U., Chaudhuri, S., Madigan, D., Eds.; Association for Computing Machinery:New York, NY, USA, 1999.
[50] Li,W.; Han, J.; Pei, J. CMAR: Accurate and e_cient classification based on multiple class-association rules. In Proceedings of the 1st IEEE International Conference on Data Mining (ICDM'01), San Jose, CA, USA, 29 November–2 December 2001; Cercone, N., Lin, T.Y.,Wu, X., Eds.; IEEE Computer Society: Washington, DC, USA, 2001; pp. 369–376.
[51] Baralis, E.; Garza, P. A lazy approach to pruning classification rules. In Proceedings of the IEEE International Conference on Data Mining, Maebashi City, Japan, 9–12 December 2002; IEEE Computer Society: Washington,DC, USA, 2002; pp. 35–42.
[52]. Baralis, E.; Chiusano, S.; Garza, P. On support thresholds in associative classification. In Proceedings of theACM Symposium on Applied Computing, Nicosia, Cyprus, 14–17 March 2004; Association for Computing Machinery: New York, NY, USA, 2004; pp. 553–558.
[53] Han, J.; Pei, J.; Yin, Y. Mining frequent patterns without candidate generation. In Proceedings of the 2000 ACM-SIGMID International Conference on Management of Data, Dallas, TX, USA, 16–18 May 2000; Association for Computing Machinery: New York, NY, USA, 16 May 2000; pp. 1–12.
[54] Quinlan, J.R.; Cameron-Jones, R.M. FOIL: A midterm report. In Proceedings of the 6th European conference on Machine Learning, Vienna, Austria, 5–7 April 1993; Brazdil, P.B., Ed.; Springer-Verlag: Berlin, Germany,1993; pp. 1–20.
[55] Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.;Witten, I.H. The WEKA Data Mining Software: An Update. SIGKDD Explor. 2009, 11, 10–18.