Tec Empresarial
P-ISSN: 1659-2395; E-ISSN: 1659-3359

BREAST CANCER DETECTION BY RULE-BASED CLASSIFICATION USING HARMONY SEARCH FROM MAMMOGRAM

Manmohan Sahoo
KMBB College of Engg, Khurda,Odisha
manu_mohan@yahoo.co.in

Amalendu Bag
Regional college of Management, Bhubaneswar,Odisha
amalendu.bag@gmail.com

Aswini Kumar Mohanty
Synergy Institute of Technology, Bhubaneswar,Odisha
asw_moh@yahoo.com

Abstract -- Breast tumour converted cancer is the leading cause of cancer death among women. Screening mammography is the reliable and low cost method currently available for detection of early and potentially curable breast cancer. Research indicates that the mortality rate could decrease by 30% to 40% if women age 50 and older have regular mammograms. The detection rate can be increased 10-20% by providing the radiologist with results from a computer-aided diagnosis (CAD) system acting as a second opinion.

However, among screening mammograms routinely interpreted by radiologists, very few (approximately 0.5%) cases actually have breast cancer. It would be beneficial if an accurate CAD system existed to identify normal mammograms and thus allowing the radiologist to focus on suspicious cases. This strategy could reduce the radiologist's workload and improve screening performance. Masses, Calcification and Architectural distortion are the three major signs notable as causes for cancer recognized using mammogram images

Image mining is concerned with knowledge discovery in image databases. Since mammography is considered as the most effective means for breast cancer diagnosis, this paper introduces multi dimensional genetic association rule mining for classification of mammograms. The approach involves Image Pre-processing resources to extract morphological features from tumours in mammograms and Image Mining to classify them as benign or malignant.

Results will show that there is promise in image mining based on multi dimensional genetic association rule mining. It is well known that data mining techniques are more suitable to larger databases than the one used for these preliminary tests. Images from MIAS repository were used for the experiments. The results showed the efficacy of the proposed method and system, which reduced the false positive and false negative rates, and allowed a more efficient decision-making process.

Keywords—Mammogram, morphological features, Classification,; Association rule mining using harmony search , Computer Aided Diagnosis, statistical features

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 [2,3].

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 [4,5], statistical methods and most of them used feature extracted using image processing techniques [6].Some other methods are based on fuzzy theory [7,8] and neural networks [9]. In this paper we have used classification method called a novel Multidimensional Genetic Association Rule Miner (MGARM) is proposed for rule construction. The result shows that the proposed rule-based approach reaches the classification accuracy over 97% and also demonstrates the use and effectiveness of association rule mining in image classification [10-13].

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 genetic association rule 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 Breast tumour morphology followed by pre-processing, segmentation. Section 3 presents the histogram equalization and feature extraction phase. Section 4 discusses the description of methodology and section 5 presents method of Feature selection and classification followed by section 6. In section 7 the results are discussed and conclusion is presented in section 8.

2. BREAST TUMOR MORPHOLOGY

The disorderly growth and multiplication of cells result in tumor formation. When malignant, the tumor is considered cancer, tends to invade adjacent tissues and therefore has an irregular shape and indistinct edge Alvarenga et al. 2003[14]. On the other hand, benign tumors tend to have

rounded or oval shape and circumscribed edge. Thus, the tumors usually are visually distinguishable according to the parameters:
Size: small tumors cannot be detected through clinical exams, and can only be identified through mammograms;
Shape: tumors may have round, oval or irregular shape;
Margin: malignant tumors generally have an irregular or undefined contour;
Density: it is possible to classify the tumor density in relation to the surrounding normal glandular tissue. Malignant tumors tend to have high density and appear in the image as white areas.

2.1. In Image Processing, techniques are applied to improve the image quality, high- lighting the edges of objects and eliminating noise acquired in image acquisition dos Santos Romualdo et al. 2009[15]. This process can include operations such as image preprocessing, identification of regions of interest, feature extraction and classification (recognition) of objects.

2.1.1 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 [16]. 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. In pre-processing phase, techniques are applied to improve image visualization and interpretation by the observer. There are many enhancement techniques, which choice and use are directly related to the context and to the analyst preference in respect to a “good image” Gonzalez and Woods 2008[17].

A pre-processing; usually noise-reducing step is applied to improve image and calcification contrast figure 2. In this work the Negative Transformation technique was used. This technique reverses the image histogram and, consequently, the grayscale of the image. It highlights white or gray details in dark regions of an image Gonzalez and Woods 2008[17].

2.1.2. Image Segmentation

Segmentation is the subdivision of an image into its constituent regions until the objects of interest are identified Gonzalez et al.[17]. This technique is very important in pattern recognition tasks Cuadros et al. [18].

Among several segmentation techniques, the region growing (RG) was used in this work due to its wide application in the literature. This method is based on the fact that an image is composed by a set of regions and there, in turn, are formed by a set of pixels. The growth results in grouped regions according to predetermined parameters such as similarity of grayscale or texture. The definition of growth parameters depends on the problem explored and on the image type. In this work the mammograms were segmented based on the gray levels.

In Figure 1(a) the mammogram can be observed. In Figure 1(b) the mammogram after the negative transformation is presented. Figure 1(c) brings the result of the image segmentation of Figure 1(b). In Figure 1(d), the ROI is selected ignoring the rest of the image.

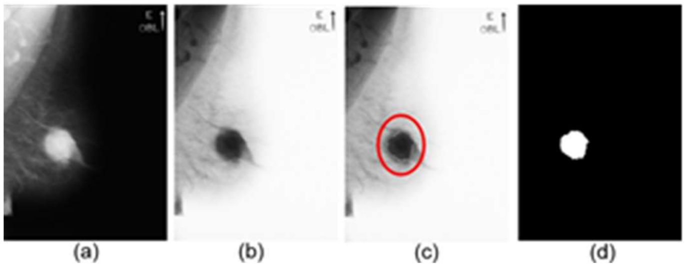


Figure 1. (a) Original image (b) Negative transformation (c) Segmented image (d) ROI

3. HISTOGRAM EQUALIZATION

Histogram equalization is a method in image processing of contrast adjustment using the image's histogram [19]. 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.

† peipa.essex.ac.uk/info/mias.html

3.1. 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 [20,21]. 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. The process of feature extraction allows obtaining relevant characteristics to represent lesions found in mammograms. There are many attributes to represent ROI’s on medical images, which are distributed in three main categories: texture, gradient and geometry (morphology) Nixon and Aguado [22].

Texture describes image characteristics such as roughness, uniformity and regularity. The recognition of texture features is a challenge because they do not present a regular pattern and are dependent on scale Ferreira et al. [23]. Texture analysis is usually a very time consuming process, so it is not indicated in scenarios where a large volume of images is involved Costa et al. [24]. The gradient is ideal to find changes in gray levels and the direction of edge on image regions Gonzalez et al. [17].

The geometric attributes describe the morphological properties of regions of interest as area, perimeter and circularity. In this work, the morphological attributes were extracted from mammograms due to their high relevance in medical diagnosis to identify objects (regions) of interest Al-Shamlan et al.[25], and because breast tumors are visually distinguishable by their geometric shape.

4. DESCRIPTION OF THE METHODOLOGY

Based on the proposal of Souto et al. [26], the methodology implemented in the CAD system, is divided into two phases: training and classification (Figure 2).

The following tasks are performed during the training phase:

Selection: the mammogram that will be submitted to the training process is selected from an image database. In this phase, the selected image already has a diagnosis;

Transformation: the image is submitted to negative transformation to highlight the lesions, helping the analyst to better observe the ROIs;

Segmentation: in this work, the segmentation performed by the region growing algorithm can be considered as semi-automatic and local. The user indicates the initial seed point of the segmentation and empirically sets the Euclidean Distance (ED) according to the characteristics of the lesion.

Feature extraction: geometric features such as area, perimeter, perimeter-area, shape and fractal (Table 1 and 2) from ROI are calculated including intensity histogram features for hoping better results due to texture;

Creation of the model: features extracted from a set of images and their corresponding diagnoses are used as input for the classification algorithm (Association rule based using harmony search);

Model: the model characterizes the king of lesion (benign or malignant) in mammograms and is used in the classification phase.

The classification phase is similar to training phase:

Selection: the mammogram that will be used in the classification process is selected. It does not have a diagnosis in this phase;

Transformation: the image is submitted to negative transformation to highlight the lesions;

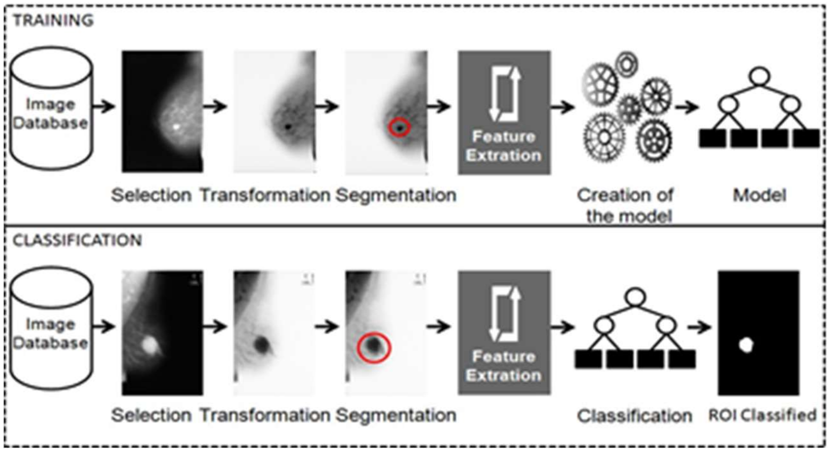
Segmentation: the identification and segmentation of the ROI are the objectives of this step, similarly to this procedure on the training phase;

Feature extraction: features (Table 1 and 2) are obtained from the ROI;

Classification: the extracted features are submitted to the created model in the training phase. This classification process indicates the diagnosis of the lesion;

ROI classified: the result of this phase is the region of interest classified as benign or malignant. This information can help the specialist providing a second opinion about the case or increasing the security of diagnosis.

Figure 2.



5. IMAGE FEATURES

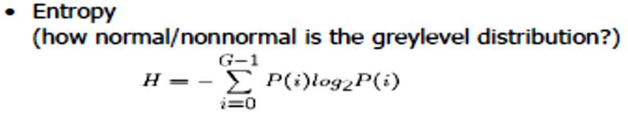
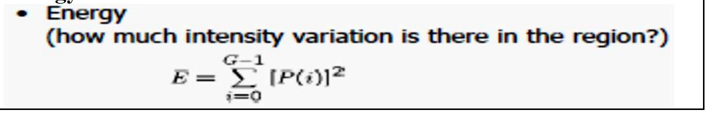
Image features in image processing are prominent areas or spots that serve as a representation of an image's visual information. They are employed to categorize, characterize, and match pictures. To perform tasks like object recognition, image segmentation, and classification, image characteristics are crucial to image processing. The ability of image features to help the computer analyze, comprehend, and interpret images makes them essential to image processing. We can describe an image's content in a way that is better suited for computer analysis by extracting features from it. Various tasks including object detection, image segmentation, and classification are made easier with the aid of this representation. Additionally, by concentrating on an image's key characteristics, image features in image processing allow us to simplify an image and make it simpler to process and analyze.

5.1. Intensity Histogram Features

Intensity Histogram analysis has been extensively researched in the initial stages of development of this algorithm [27]. Prior studies have yielded the intensity histogram features like mean, variance, entropy etc. These are summarized in Table I Mean values characterize individual calcifications; Standard Deviations (SD) characterize the cluster.

TABLE 1
INTENSITY HISTOGRAM FEATURES

Feature Number assigned	Feature
1.	$\mu = \frac{\sum_{i=0}^{G-1} ip(i)}{\sum_{i=0}^{G-1} p(i)} = \frac{\sum_{i=0}^{G-1} ip(i)}{n} = \sum_{i=0}^{G-1} iP(i)$ Mean
2.	Variance • Variance (a more credible feature, measures region "roughness") $\sigma^2 = \sum_{i=0}^{G-1} (i-\mu)^2 P(i)$
3.	Skewness • Skewness (are the texel intensities usually darker/lighter than average?) $m_3 = \sum_{i=0}^{G-1} (i-\mu)^3 P(i)$
4.	• Kurtosis (how "uniform" is the greylevel distribution?) $m_4 = \sum_{i=0}^{G-1} (i-\mu)^4 P(i)$ Kurtosis

5.	**Entropy**  • Entropy (how normal/nonnormal is the greylevel distribution?) $H = -\sum_{i=0}^{G-1} P(i)log_2P(i)$
6.	**Energy**  • Energy (how much intensity variation is there in the region?) $E = \sum_{i=0}^{G-1} [P(i)]^2$

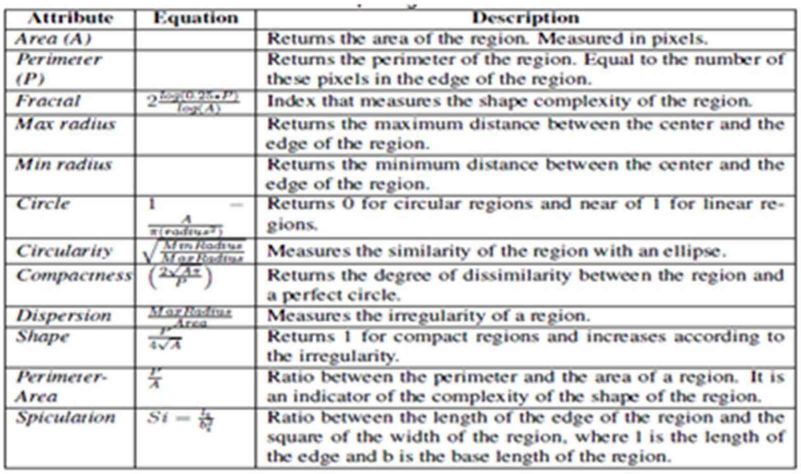
5.2. Morphological properties for feature extraction of geometrical shapes

Geometric shapes are very important in the field of pattern recognition and have played a large role in the field of computer vision and have adopted many algorithms in this field. When studying the features of the co-occurrence matrix the power of certain features and the weakness of others appear. In the shapes, the features of homogeneity, energy, and contrast were of high affectivity especially when eliciting the characteristic of the geometrical shapes that have been counted after certain primary processing and operating of cutting and reshaping the structure needed for the shape.

Detection of geometric features in digital images like mammograms is more important exercise in image analysis and computer vision [28]. Many properties of objects in our world are strongly determined by geometric properties, the applications of shape analysis extend over almost every applied scientific and technological area, from the smallest to the largest spatial scales. For instance, the strength of composites is directly related to the shape of its constituent crystal grains, the shape of biological entities provide an immensely important clue about the interactions between themselves and with the environment. The relationship is no less dramatic at larger scales, where the properties of a wing design or a mechanical piece are almost completely defined by their respective geometry. Indeed, when properly and carefully applied shape analysis provide an exceedingly rich potential for applications in the most diverse areas, from material sciences to biology and neuroscience [29]. The Hough Transform (HT) is a standard method for shape recognition in digital images [30,31]. It was initially used to recognition straight lines [32,33] and later extended to circle [34], ellipses [35] and arbitrarily shaped objects [36]. Its advantages include robustness to noise, robustness to shape distortions and to occlusions/missing parts of an object. Its main disadvantage is the fact that computational and storage requirements of the algorithm increase as a power of the dimensionality of the curve. This means that for in a straight line the computational difficulty and storage space supplies are O(n2), for circles O(n3) and for ellipses O(n5[37].

TABLE II

Morphological attributes



Attribute	Equation	Description
Area (A)		Returns the area of the region. Measured in pixels.
Perimeter (P)		Returns the perimeter of the region. Equal to the number of these pixels in the edge of the region.
Fractal	$2^{\frac{\log(0.25*P)}{\log(A)}}$	Index that measures the shape complexity of the region.
Max radius		Returns the maximum distance between the center and the edge of the region.
Min radius		Returns the minimum distance between the center and the edge of the region.
Circle	$1 - \frac{A}{\pi(radius^2)}$	Returns 0 for circular regions and near of 1 for linear regions.
Circularity	$\sqrt{\frac{MinRadius}{MaxRadius}}$	Measures the similarity of the region with an ellipse.
Compactness	$\left(\frac{2\sqrt{A\pi}}{P}\right)$	Returns the degree of dissimilarity between the region and a perfect circle.
Dispersion	$\frac{MaxRadius}{Area}$	Measures the irregularity of a region.
Shape	$\frac{P}{4\sqrt{A}}$	Returns 1 for compact regions and increases according to the irregularity.
Perimeter-Area	$\frac{P}{A}$	Ratio between the perimeter and the area of a region. It is an indicator of the complexity of the shape of the region.
Spiculation	$Si = \frac{l}{b^2}$	Ratio between the length of the edge of the region and the square of the width of the region, where l is the length of the edge and b is the base length of the region.

6. CLASSIFICATION

6.1. Apriori algorithm and interesting measures

Apriori is a standard and well-known basic algorithm in association rule mining that is used for mining frequent itemsets in a set of transactions. It was first introduced by Agrawal and Srikant [8]. The APRIORI-C is another Apriori-based algorithm that drives rules according to the parameters minimal confidence and minimal support of a rule Jovanoski & Lavrač, [39]. Predictive Apriori Scheffer, [40] in another algorithm motivated by Apriori and unlike the confidence related focus of Apriori tries to maximizes the expected accuracy of an association rule on unseen data. While Apriori sorts the rules based on confidence only, Predictive Apriori considers both the confidence and support in ranking the rules.

Nahar et al. considered three rule generation algorithms—Apriori, Predictive Apriori and Tertius-for extraction the meaningful factors for particular types of cancer Nahar et al., [41] and heart disease Nahar et al., [42]. Their experimental results showed that Apriori is the most beneficial association rule mining algorithm.

$$Support\,(X \Longrightarrow Y) = P(X \cap Y)$$
$$Confidence\,(X \Longrightarrow Y) = P(Y|X) = \frac{Support\,(X \cap Y)}{Support\,(X)}$$
$$Lift\,(X, Y) = \frac{P(X \cap Y)}{P(x) * P(Y)}.$$

Apriori algorithm can produce a lot of rules, but much of them are superfluous. To select appropriate rules from the set of all possible rules, constraints on various measures of interestingness can be used. Support and confidence are two measures of rule interestingness that mirror the usefulness and certainty of a rule respectively Agrawal et al., [43]. The support is the percentage of the total number of records of transactions that include all items in the antecedent (if) and consequent (then) parts of the rule. Frequent itemsets are those itemsets that their frequency is greater than a predefined minimum support (Minsup). Confidence is the ratio of the number of transactions that include all items in the consequent, as well as the antecedent (the support) to the number of transactions that include all items in the antecedent. In other words, confidence is the accuracy of the rule and usually is used in Apriori for ranking the rules. The task of association rule mining is to generate all association rules from the set of transactions that have a support greater than Minsup and confidence greater than Mincon. Since we need to discover the relationshipbetween input attributes and class label, we need to find all the rules of the form A → B that antecedent part of the rule includes of some item and the consequent part can just be the class items.

High support and high confidence rules are not necessarily interesting. Instead of using only support and confidence, we also used lift measure as a metric for evaluating the significance and reliability of association rules. Lift is the ratio of Confidence to Expected Confidence. Hence, Lift is a value that gives us information about the increase in the probability of the consequent given antecedent part of a rule. A lift ratio larger than1.0 implies that the relationship between the antecedent and the consequent is more significant than would be expected and make those rules potentially useful for predicting the consequent in unseen instances. The larger the lift ratio, the more significant the association.

Another issue that must be considered is related to the type of dataset that is appropriate for applying the Apriori algorithm. Consider a dataset for supervised learning which contains observations of a class label variable and a number of predictor variables. Such a dataset can be converted into an appropriate format for association rule mining if both the class label and the predictors are of the categorical type. Since our benchmark datasets contain continuous variables, we must use a method for handling numeric attributes. There are some methods for this purpose. A traditional method is discretization that can be static or based on the distribution of data. We used a method proposed by Tsai et al.[44].

6.2. Associative rules for classification

In recent years, some researchers tried to combine association rule mining and classification [45-50]. Their experiments show that this approach achieves better accuracy than conventional classification algorithms such as C4.5. The reason is that the associative classifier is composed of high-quality rules, which are generated from highly confident event associations that reflect the close dependencies among events.

The Classification Based on Association rules (CBA) algorithm is one of the first efforts for combining of classification and association rule mining (Ma et. al[47]. This algorithm will describe with details in the next section. Li, Han et. al[46] suggested a weighted $\chi 2$ analysis to perform a Classification based on Multiple Association Rules(CMAR). Unlike the CBA algorithm, the CMAR algorithm uses all the rules that cover the example to be classified instead of using just one rule.

Yin & Han [50] propose the CPAR (Classification based on Predictive Association Rules) rule-based classification algorithm CPAR doesn’t generate a large number of candidate rules as in conventional associative classification. It pursues a greedy algorithm to produce rules directly from training data and uses the best K rules in prediction time.

An advantage of associative classifiers is that they are rule-based and thus lend themselves to be more easily understood by humans. As previously stated, a classification system is built in two phase. In the first stage, the learning target is to discover the association patterns inherent in a database (also referred to as knowledge discovery). In the second stage, the goal is to select a small set of relevant association patterns to construct a classifier given the predictor attributes. To produce the best classifier out of the entire set of rules, we need to consider all the feasible subsets of rules and selecting the most accurate subset. This is clearly impractical.

In the classification phase, some methods such as (Ma et al.[47]; Thabtah et al.[51]; Wanget al. [52]), simply select a rule with a maximal user-defined measure, such as confidence. If there is no rule covering the example, then usually the prevalent class is taken to be the predicted class. However, identifying the most effective rule at classifying a new case is a big challenge. When classifying a new data object, it may have more rules that satisfy the test conditions and using them may increase the prediction accurately Li, Han et al.[46].

6.3. CBA algorithm

Classification Based on Associations (CBA) algorithm is one of the first algorithms to bring up the idea of classification using association rules Ma & Liu et al [47]. CBA implements the famous Apriori algorithm (Agrawal et al. [38]) in order to discover frequent items. Once the discovery of frequent items finished, CBA proceeds by converting any frequent item that passes the Minconf into a rule in the classifier. At the rule generation phase, CBA selects a special subset of association rules whose right-hand-side are restricted to the classification class attribute. This subset of rules is called class association rules (CARs). At the next step, the CBA algorithm builds a classifier using CARs. At this step, CBA uses a heuristic approach and sorts the rules according to their confidence and selects top rules that cover the training samples.

The algorithm first selects the best rule (rule having the highest confidence), then eliminates all the covered examples. If at least one example satisfied the rule conditions, then that rule is appended to the final rules. This procedure is repeated until there are no more rules to select or there are no more examples to cover. The algorithm then stops and returns the classifier in the form of an IF-THEN-ELSE rule list. One challenge with this approach is that selecting the best rules may be not the best subset of rules.

The CBA system follows the original association rule model and uses a single Minsup in its rule generation. It seems that this is inadequate for mining of CARs because class frequency distributions in many practical classification datasets are unbalanced. We used the CBA algorithm with three little changes. The first change is that we use multiple Minsup than can be useful for imbalanced datasets. The second change is that in the original CBA algorithm once each sample is covered by a rule, it is removed from the samples; we defined a parameter called Delta. This parameter defined that how many times each sample must be covered to remove from the samples Li, Han & Pei, 2001[46]). This approach leads to the generation of more rules. The third change occurs in the classification phase. In the classification phase of the original CBA algorithm, the rule with maximum confidence that covers the test conditions defines the class label of a test sample. We select the top K (a predefined parameter) rules from each class that covers the test sample conditions and determined the class label according to the sum of the confidence of selected rules.

All data preprocessing and analyse were conducted using Matlab as well as WEKA.

6.4. Proposed method using harmony search

The proposed method of rule selection based on HS is depicted in Fig. 1. At the initial step, we did some preprocessing on each dataset. One of the main preprocessing is discretization of continuous features. We applied a discretization algorithm based on a class-attribute contingency coefficient that was proposed by Tsai et al.[53]. After discretization, we convert each dataset to the appropriate format such that the value of each feature can be True (1) or False (0). For this aim, if a feature is converted to N different discrete values, we produce N feature. After the conditions are satisfied for the Apriori algorithm, we run this algorithm for each class with different Minsup and Minconf. The main novelty of our study is in the next step. As previously was mentioned, the Apriori algorithm produces many rules and CBA algorithm uses a greedy algorithm for selecting a subset of produced rules for building a classifier. Using greedy approaches cause the selected rules to not be the best subset of rules.

We believe that population-based evolutionary algorithms fit well to the rule selection problem. Harmony Search (HS) is a population-based stochastic search algorithm that inspired by the musical process of searching for a perfect state of harmony Geem, Kim et al.[54]. The harmony in music is analogous to the optimization solution vector, and the musician's improvisations are similar to local and global search methods in optimization techniques When a musician is improvising, he has three choices: (1) to execute any pitch from memory; (2) to execute a pitch next to any other in his memory;(3) to execute a random pitch from the range of all possible pitches. These three options are employed in the HS algorithm by means of three main parameters: Harmony Memory (HM), Harmony Memory Consideration Rate (HMCR), and Pitch Adjustment Rate (PAR). The HMCR is defined as the probability of selecting a component from the present HM members. The PAR determines the probability of a candidate from the HM to be mutated. The steps in the procedure of HS are as follows:

Step 1. Initialize a harmony memory (HM). The initial HM consists of a given number of randomly generated solutions to the optimization problems under consideration.

Step 2. Improvise a new harmony from HM.

Step 3. Update the HM. If the New Harmony is better than the worst harmony in HM, then include the New Harmony in the HM and exclude the worst harmony from the HM.

Step 4. If the stopping criteria we not satisfied, go to step 2.

HS has been successfully applied to various discrete optimization problems such as Maximum Clique Problem Afkhami et al. [55]), traveling salesperson problem Geem et al. [54]), tour routing Geem et al.[56],water network design Geem [62], dynamic relocation of mobile base stations in wireless sensor networks Moh'd Alia, 2017[57], and others.

In binary HS, the size of each solution equals the number of candidate's rules. For example, if the Apriori algorithm produces 100 rules that satisfy Minsup and Minconf conditions then the size of each solution in HS will be equal to 100. Each solution consists of a binary vector of rule incidences, indicating exclusion (0) or inclusion (1) of the rule in the combination.

The standard Harmony Search (HS) is not suitable for binary representations. This is due to the pitch adjusting operator not being able to perform the local search in the binary space. Therefore we used the implementation of HS that proposed by Afkhami et al. [56]).

We run the HS algorithm with the following parameters: maximum number of iterations = 20, harmony memory size = 100, Number of new harmonies = 20, harmony memory consideration rate = 0.75.

We used Harmony Search, a music-inspired stochastic search algorithm, for selecting the best subset of rules as a classifier. One of the important sections in any meta-heuristic algorithm is the calculation of cost function. For this aim, we apply a modified version of the CBA algorithm on the selected rules and calculate the error rate of applying the resulted rules on the training and validation data. At final, the solution with the minimum cost value is selected and this solution (a subset of rules) applies on the test data. It is obvious that the proposed flowchart is shown for one fold of cross-validation. In K-fold cross-validation, this approach must be repeated for K times, until all the samples in the dataset are used for the test data. The pseudo code of the proposed method is shown in Table 1.

Time complexity of the Apriori algorithm and association rule mining is a critical challenge that must be considered (Cano et al.[45]; Zafra et al.[58]; Luna et al., 2016[59]; Thabtah et al.[60]). As its time complexity is exponential, we can do some preprocessing activity to decrease the running time. First of all, we can apply feature selection before applying Apriori algorithm. Feature selection can be done before or after of discretization. The second thing that we can do is related to the size of the rules. As small rules are favorable, we can limit the size of items that appear in a rule and consequently decrease the running time of Apriori algorithm.

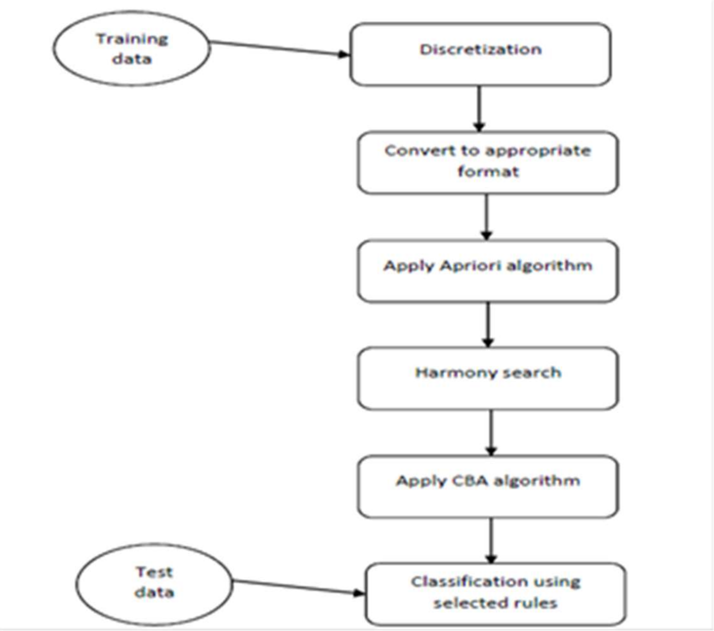


Figure 1 The framework of the proposed method. This figure shows what steps must be done for implementation of the proposed method.

Table 1 Pseudo code of the proposed method. This pseudo code supposes that we have training input, training output, test input, validation input and validation output. The code shows that how we build a rule-based classifier and determine the test data output.

```
For $i = 1$ to K fold
Determine Traininput, Trainoutput, Testinput, Testoutput, Valinput and
Valoutput  Finalrules={};
For $j = 1$ to number_class
Rulesj= apply Apriori algorithm(traininput, Minsupj,Minconj,class j )
Finalrules= append Rulesj to Finalrules
End %for j
Selected_rules=Apply harmony search algorithm (Finalrules, Traininput, Trtainoutput, Valinput,
Valoutput)
     Testoutput= apply selected_rules on Testinput
          End %for i
```

7. EXPERIMENTAL RESULTS

The digital mammograms used in our experiments were taken from the Mammographic Image Analysis Society (MIAS). The database consists of 322 images, which belong to three categories: normal, benign and malign (ftp://peipa.essex.ac.uk). There are 209 normal images, 62 benign and 51 malign, which are considered abnormal.

The proposed method is evaluated based on ten-fold cross validation method. The following table presents the rule accuracy of the proposed classification system compared with other association rule based system proposed in [40, 41]. The results for the ten splits of the mammogram database are given in Table IV.

TABLE IV
CLASSIFICATION ACCURACY FOR THE TEN SPLITS WITH MGARM

Splits Classification Accuracy

Splits	Classification Accuracy
1	98.73
2	97.97
3	97.84
4	95.20
5	93.31
6	98.74
7	95.49
8	98.62
9	98.13
10	96.97
Average	**97.20**

In this paper we used multi dimensional genetic association rule mining using image contents for the classification of mammograms. The average accuracy is 97.20 %. We have employed the freely available Machine Learning package, WEKA [61]. Out of 322 images in the dataset, 230 were used for training and the remaining 92 for testing purposes and the result is shown in Table V.

TABLE V
RESULTS OBTAINED BY PROPOSED METHOD

Normal	97.60%
Malignant	92.15%
Benign	100%

The confusion matrix has been obtained from the testing part .In this case for example out of 51 actual malignant images 06 images was classified as normal. In case of benign all images are correctly classified and in case of normal images 6 images are classified as malignant. The confusion matrix is given in Table VI.

TABLE VI
CONFUSION MATRIX

Actual	**Predicted class**		
	Benign	**Malignant**	**Normal**
Benign	63	0	0
Malignant	51	47	04

Normal	209	5	204

8. CONCLUSIONS

Computer based breast cancer detection has been studied for more than two decades Mammography is one of the low cost and reliable methods in breast cancer detection, but in some cases radiologists face difficulty in differentiating the tumors whether benign or malignant causing patients panic and mental agony and referred to biopsy . We have described here a comprehensive technique for quick and accurate detection whether it is malignant or benign from the tumor which require less efforts and in quick decision with methods in a uniform terminology, to define general properties and requirements of local techniques as well as to enable the readers to select the efficient method that is optimal for the specific application in detection of micro calcifications in mammogram images.

Classification of Microcalcification Clusters (MCs) is one of the key points to find the early sign of breast cancer. In this paper, we have proposed a novel association rule based system for classification of Microcalcification Clusters (MCs). Initially the MCs are segmented from the mammograms and the statistical GLCM features including geometrical/morphological features are extracted to enable the better classification and to avoid the classifier from over fitting either or under fitting which can affect the accuracy level. The proposed approach for associative classifier with harmonic search clearly outperform the other classifiers as the result shown. The proposed and described classifier is applied to construct the association rule to classify the images into three classes: normal, benign and malign. The result shows that the accuracy level is somewhat justifiable. it is quite noticeable that combining both texture features as well as morphological features have high impact rather than a single type of feature. Suitable feature subset selection may improve the accuracy level might improve the level of accuracy. In future, an efficient algorithm can be used to select the relevant features including glcm, glrlm, features would be taken into account and the rules can be generated to improve the 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 Expectation
Maximum(EM) algorithm for mammography" , International Journal of Imaging System and
Technology, 8(5): pp491-504,1997.

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

[6] Jelena Bozek, Mario Mustra, Kresimir Delac, and Mislav Grgic "A Survey of Image Processing Algorithms in Digital mammography"Grgic et al. (Eds.): Rec. Advan. in Mult. Sig. Process. and Commun., SCI 231, pp. 631–657,2009

[7] 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.

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

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

[10] 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

[11] 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.

[12] 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.

[13] 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

[14] Alvarenga, A. V., Infantosi, A. F. C., de Azevedo, C. M., and de Albuquerque Pereira,W. C. (2003). Application of morphological operators in the segmentation and determination of the contour of breast tumors in ultrasound images. Brazilian Journal of Biomedical Engineering, 19(2):91–101.

[15] dos Santos Romualdo, L. C., da Costa Vieira, M. A., and Schiabel, H. (2009). Mammography images restoration by quantum noise reduction and inverse mtf filtering. SIBGRAPI,22th Brazilian Symposium on Computer Graphics and Image Processing:180–185.

[16] Pisano ED, Gatsonis C, Hendrick E et al. “Diagnostic performance of digital versus film mammography for breast-cancer screening”. NEngl J Med 2005; 353(17):1773-83.

[17] Gonzalez, R. C. and Woods, R. E. (2008). Digital Image Processing. Prentice Hall, 5 edition.

[18] Cuadros, O., Botelho, G., Rodrigues, F., and Neto, J. B. (2012). Segmentation of large images with complex networks. SIBGRAPI, 25th Brazilian Symposium on Computer Graphics and Image Processing:24–31.

[19] Wanga X, Wong BS, Guan TC. ‘Image enhancement for radiography inspection”. International Conference on Experimental Mechanics. 2004: 462-8.

[20] 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

[21] 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.

[22] Nixon, M. S. and Aguado, A. S. (2008). Feature Extraction & Image Processing. Academic Press, 2 edition.

[23] Ferreira, C. B. R. and Borges, D. L. (2005). A selection strategy of a minimal subset of wavelet features in a multiresolution approach for the classification of tumors on mammograms. V Workshop of Medical Informatics, 1.

[24] Costa, A. F., Humpire-Mamani, G., and Traina, A. J. M. (2012). An efficient algorithm for fractal analysis of textures. SIBGRAPI, 25th Brazilian Symposium on Computer Graphics and Image Processing:39–46.

[25] Al-Shamlan, H. and El-Zaart, A. (2010). Feature extraction values for breast cancer mammography images. International Conference on Bioinformatics and Biomedical Technology, pages 335–340.

[26] Souto, L. P. M., dos Santos, T. K. L., and dos Santos Silva, M. P. (2014). Metricas morfolgicas para a classificacao detumores de mama. Workshop de Informatica Medica.CSBC.

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

[28] Yu,Wing Hong W. 1995. Memory –Efficient Circle and Ellipse Detection Algorithm in Digital Images, Dept.of Industrial and Manufacturing Systems Engineering , university of Hong Kong, IEEE International conference, Volume 44, ,Pages 4262 - 4267.

[29] Roberto M. Cesar Junior, Luciano da Fontoura Costa 2001. Shapes Analysis And Classification, Baco Raton, FL, USA, Pages 1 - 5.

[30] V.F. Leavers 1993. Survey: which Hough transform, Computer Vision, Graphics and Image Processing: Image Understanding, Volume 58, Issue 2, Pages 250-264.

[31] J. Illingworth, J. Kittler 1988. A Survey Of The Hough Transform, Computer Vision, Graphics and Image Processing, Volume 44, Issue 1, Pages 87–116.

[32] A. Rosenfeld 1969. Picture Processing by Computer, Academic Press, New York.

[33] R.O. Duda, P.E. Hart 1972. Use of the Hough Transformation to detect lines and curves in pictures, Communications of the Association of Computing Machinery, Academic Journal, Volume 15, Issue 1, Pages 11–15.

[34] E.R. Davies 1987. A modified Hough scheme for general circle location, Journal Pattern Recognition, Elsevier Science Inc. New York, NY, USA, Volume 7, Pages 37– 43.

[35] R.K.K. Yip, P.K.S. Tam, D.N.K. Leung 1995. Modification of Hough transform for circles and ellipses detection using a 2-dimensional array, Pattern Recognition , Volume 25, No. 11, Pages 1733–1744.

[36] D.C.W. Pao, H.F. Li, R. Jayakumar 1992. Shapes recognition using the straight line Hough transform: theory and generalization, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 14, Issue 11 ,Pages 1076–1089.

[37] D.Loannou, W.Huda, A. F. Laine 1999. Circle Recognition Through a 2D Hough Transform and Radius Histogramming, Image and Vision Computing, Volume17, Issue 1, page 15-26.

[38] Agrawal R, Imieliński T, Swami A. 1993. Mining association rules between sets of items in large databases. In: Proceedings of the 1993 ACM SIGMOD international conference on management of data, Washington DC, 25-28 May 1993. New York: ACM, 207–216 DOI 10.1145/170035.170072

[39] Jovanoski V, Lavrač N. 2001. Classification rule learning with APRIORI-C. In: Portuguese conference on artificial intelligence. Springer, 44–51.

[40] Scheffer T. 2001b. Finding association rules that trade support optimally against confidence. Principles of Data Mining and Knowledge Discovery 424–435 DOI 10.1007/3-540-44794-6_35.

[41] Nahar J, Tickle KS, Ali AS, Chen Y-PP. 2011. Significant cancer prevention factor extraction: an association rule discovery approach. Journal of Medical Systems 35:353–367 DOI 10.1007/s10916-009-9372-8.

[42] Nahar J, Imam T, Tickle KS, Chen Y-PP. 2013. Association rule mining to detect factors which contribute to heart disease in males and females. Expert Systems with Applications 40:1086–1093 DOI 10.1016/j.eswa.2012.08.028.

[43] Agrawal R, Mannila H, Srikant R, Toivonen H, Verkamo AI. 1996. Fast discovery of association rules. Advances in Knowledge Discovery and Data Mining 12:307–328.

[44] Tsai C-J, Lee C-I, Yang W-P. 2008. A discretization algorithm based on class-attribute contingency coefficient. Information Sciences 178:714–731 DOI 10.1016/j.ins.2007.09.004

[45] Cano A, Zafra A, Ventura S. 2013. An interpretable classification rule mining algorithm. Information Sciences 240:1–20 DOI 10.1016/j.ins.2013.03.038.

[46] Li W, Han J, Pei J. 2001. CMAR: accurate and efficient classification based on multiple class-association rules. In: Data mining, 2001 ICDM 2001, proceedings IEEE interna- tional conference on. Piscataway: IEEE, 369–376.

[47] Ma BLWHY, Liu B. 1998. Integrating classification and association rule mining. In: Proceedings of the fourth international conference on knowledge discovery and data mining.

[48] Wang K, Zhou S, He Y. 2000. Growing decision trees on support-less association rules. In: Proceedings of the sixth ACM SIGKDD international conference on knowledge discovery and data mining. New York: ACM, 265–269.

[49] Wang Y, Wong AKC. 2003. From association to classification: inference using weight of evidence. IEEE Transactions on Knowledge and Data Engineering 15:764–767 DOI 10.1109/TKDE.2003.1198405.

[50] Yin X, Han J. 2003. CPAR: classification based on predictive association rules. In: Proceedings of the 2003 SIAM international conference on data mining. Philadelphia: SIAM, 331–335

[51] Thabtah FA, Cowling P, Peng Y. 2004. MMAC: a new multi-class, multi-label associative classification approach. In: Data mining, 2004 ICDM'04 fourth IEEE international conference on. Piscataway: IEEE, 217–224.

[52] Wang K, Zhou S, He Y. 2000. Growing decision trees on support-less association rules. In: Proceedings of the sixth ACM SIGKDD international conference on knowledge discovery and data mining. New York: ACM, 265–269.

[53] Tsai C-J, Lee C-I, Yang W-P. 2008. A discretization algorithm based on class-attribute contingency coefficient. Information Sciences 178:714–731DOI 10.1016/j.ins.2007.09.004.

[54] Geem ZW, Kim JH, Loganathan GV. 2001. A new heuristic optimization algorithm:harmony search. Simulation 76:60–68 DOI 10.1177/003754970107600201.

[55] Afkhami S, Ma OR, Soleimani A. 2013. A binary Harmony Search algorithm for solving the maximum clique problem. International Journal of Computer Applications 69(12):38–43.

[56] Geem ZW, Tseng C-L, Park Y. 2005. Harmony search for generalized orienteering problem: best touring in China. In: International conference on natural computation. Springer, 741–750.

[57] Moh'd Alia O. 2017. Dynamic relocation of mobile base station in wireless sensor networks using a cluster-based harmony search algorithm. Information Sciences 385:76–95.

[58] Cano A, Zafra A, Ventura S. 2014. Parallel evaluation of Pittsburgh rule-based classifiers on GPUs. Neurocomputing 126:45–57 DOI 10.1016/j.neucom.2013.01.049.

[59] Luna JM, Cano A, Pechenizkiy M, Ventura S. 2016. Speeding-up association rule mining with inverted index compression. IEEE Transactions on Cybernetics 46:3059–3072 DOI 10.1109/TCYB.2015.2496175

[60] Thabtah F, Cowling P, Hammoud S. 2006. Improving rule sorting, predictive accuracy and training time in associative classification. Expert Systems with Applications 31:414–426 DOI 10.1016/j.eswa.2005.09.039.

[61] 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.

[62] Geem ZW. 2006. Optimal cost design of water distribution networks using harmony search. Engineering Optimization 38:259–277 DOI 10.1080/03052150500467430