Optimization of Association Rule Mining for Mammogram Classification

Poonam Sonar   *poonam.sonar@mctrgit.ac.in*
Research Scholar,
Department of Electronics and Telecommunication Engineering,
Rajiv Gandhi Institute of Technology, Mumbai.
University of Mumbai, India

Udhav Bhosle   *udhav.bhosle@mctrgit.ac.in*
Department of Electronics and Telecommunication Engineering,
Rajiv Gandhi Institute of Technology, Mumbai,
University of Mumbai, India

Abstract

Authors presents concept of image mining, an extension of data mining, for discovering image data relationship from a large collection of mammograms images. Association rule mining is the process of discovering useful and interesting rules from large datasets based on user specified minimum support and confidence values. The set of all possible item sets grows exponentially with the number of items in the database. These constraints lead to exponential search space and dataset dependent minimum support and confidence values. It generates a huge number of unnecessary rules from frequent item sets and results in weak mining performance. The authors propose two association rule optimization techniques for overcoming these problems. The first graph theory approach (OARGT) is based on objective function such that graph generated by the optimized rules is a simple graph with simple walk. The second approach is based on Multi-Objective Genetic algorithm (MOGA) with adaptive crossover and mutations (MOGAACM). Traditionally, input to ARM classifier is in binary format. The proposed system uses variable feature quantization and relevant feature selection. In MOAGAACM, ranks are assigned to rules as per fitness function. The rules with highest rank, low crossover and mutation rates are assigned and vice versa. Experimental results show that, MOGAACM generates more effective and strong association rules compared with objective function using graph theory and achieves 89.08% and 43.60 % reduction in association rules for benign and malignant class respectively for MIAS database and 80.13 % and 79.60 % reduction in association rules for benign and malignant class respectively for DDSM medical image database. Authors propose class identification using Strength of Classification Algorithm (CISCA) for classification of mammogram image into benign and malignant classes. The classification accuracy measures reported are 91.66 % for MIAS database using graph theory and 95.45 % and 92.5 % for MIAS and DDSM respectively using MOGAACM**.**

Keywords: Association Rule Mining, Support, Confidence, Multi Objective Fitness Function, Adaptive Mutation, Crossover, Graph Theory, Genetic Algorithm.

1. INTRODUCTION

In the recent years, especially in developed countries, there is rapid increase in breast cancer cases in women. Physicians use mammograms for early detection of breast cancer, as the method is more feasible and reliable. Thus, hospitals generate large numbers of mammograms every day. Therefore, physicians require an accurate and automatic reading of digital mammograms due to its huge volume. It is very critical and difficult to analyze and diagnose mammograms. So, computer aided diagnosis (CAD) systems are developed to help the physicians. In addition, building computer aided systems to assist medical staff in hospitals is gaining high importance and priority for many researchers.

Image mining system can review semantically meaningful information contained in huge amount of data for human users. It deals with extraction of hidden information, significant knowledge, image patterns and relationships [1]. Association rule mining is used for extraction of association

rules among the given data set, has received significant research interest which resulted in development of various algorithms for extraction of association rules. Association rule mining, first time discussed by Agrawal et al. [2]. Mammogram images show breast tissue   density variations, which can be used for mammogram analysis and classification. These variations can be studied by extracting the textual features of mammogram [3]. Mammogram classification based on association rule mining using low level features is discussed in [4]. They used texture features and opinion from physicians or radiologists for frequent pattern search for improving the classification accuracy. Aswini et al. [5] discussed image mining techniques using statistical and GLCM features to classify mammogram images into normal, benign and malignant class. Maria-Luiza Antonie et al. [6, 7] proposed a mammogram classification using association rules and neural network classifier. Jawed Nagi et al [7] used seeded region growing algorithm for automated breast profile segmentation. Dhawan et al. [8] used wavelet decomposition and grey level image structure features for classification of mammograms. Prathibha et al [9] used multi scale wavelet transformation for extraction of texture features from the mammograms images.

Most of the existing association rule mining algorithms are based on the support –confidence framework. The set of all possible item sets grows exponentially with the number of items in the database. The exponential growth profoundly affects the performance of association rule mining as it generates a huge number of unnecessary rules from frequent item sets. So, there is need of optimization technique to optimize the large number of association rules. Recently, large numbers of research papers have used evolutionary algorithms for mining association rules. These studies found that evolutionary algorithms particularly genetic algorithms approaches are efficient tools especially when search space is too large to use deterministic search methods [15]. Because of inherent parallel structure, GA based methods are effective for automatic processing of large amount of data and discovering meaningful and significant information. Recently, GA based approaches named ARMGA [11] uses conditional probability as a fitness function to extract high quality Boolean association rules (BAR) without generating unnecessary rules. Most of the GA based approaches uses only one or two evaluation criteria most especially confidence factor or predictive accuracy on fitness function, to extract optimized rules. This criterion evaluates the rule depending on the number of occurrences of the rule in the entire database. More the number of occurrences better is the rule. The generated rule may have many attributes involved in the rule thereby making it difficult to understand. It would be better for the user if the algorithms can generate some of the rules that are hidden inside the data. These algorithms do not give importance toward the rare events i.e. interesting rules [11, 13].

These days, some researchers [13,16,18] have framed association rule mining problems as a multi objective problem rather than mono objective problem and jointly optimize different measures, in order to extract the set of rules which are easy compared with mono-objective algorithms and mine high quality rules from the dataset with quantitative or numerical values. Ashish Ghosh et al. [10] has framed the association rule mining problem as a multi objective problem rather than mono objective problem and jointly optimized different measures to extract a set of rules. Motivated by the features of multi objective approaches, the authors propose Multi objective Genetic algorithm with adaptive mutation and crossover (MOGAACM) algorithm which jointly optimize four objectives: lift, interestingness, predictiveness and Comprehensibility. The generated rules are easy to understand, interesting and having a good trade off among the number of rules, support and confidence and another objective of the dataset. To accomplish this, our approach extends the recent GA to perform an evolutionary learning and conditional selection and maximizes four objectives. Moreover, this approach introduces re-initialization process and an adaptive crossover and mutation method to increase the diversity in the population.

The proposed method works as follows: the first step is pre-processing of mammograms, segmentation and extraction of ROI. The second step comprises features extraction from segmented region of interest (ROI) and feature vector generation. The third step consists of formation of transaction database which is given as input to Apriori algorithm for generating association rules. In the fourth step, association rules are optimized using graph theory based objective function and MOGAACM. The first graph theory approach (OARGT) based on objective

function such that graph generated by the optimized rules is a simple graph with simple walk. To accomplish this, our method extends MOGA for performing evolutionary learning with re initialization process along with adaptive crossover and mutation. Here, rank is assigned to rules as per the fitness function. The rules having highest ranks, low crossover and mutation rates are assigned and vice versa. Optimized association rules are used for effective classification of mammogram images. Authors propose Class Identification using Strength of Classification Algorithm (CISCA) for classification of mammogram image into benign and malignant classes.

The rest of the paper is organized as follows. Section 2 discusses preliminaries, Section 3 block diagram of proposed system. It consists of mammograms preprocessing, ROI detection and use of Apriori algorithms to find association rule. Section 4 discusses optimization of association of rules based on graph. Section 5 presents proposed MOGAACM algorithms and Section 6 presents classification algorithms and Section 7 presents results and discussion on classification MIAS and DDSM data base and conclusion are reported in Section 8.

2. PRELIMINARIES: ASSOCITATION RULE MINING

Initially association rules were used in market–basket analysis but these days its applications have extended to different real-world fields including E-commerce, Telecommunications, Bioinformatics, Signal Processing, etc. Association rule mining is one of the important tasks of data mining intended toward decision support. Basically, it is the process of finding some relations among the attributes /attribute values of a huge data base. Discovery of such relationship among various attributes will help in taking specific decisions. These relations can be represented as an IF-THEN statement. IF < some conditions are satisfied > THEN, < predict some values of another attribute(s) >. The condition associated in IF part is termed as Antecedent and those with THEN part is called Consequent. Association rules search for attributes that frequently occur in given dataset based on co-occurrence of those data items [7].

Let I = {I$_1$,…,I$_n$} be a set of items. D is a set of data cases. An association rule is represented as A→B, where A and B are item sets. The strength of an association rule is measured in terms of its support and confidence. Support value gives how frequently a rule is applicable for a given data set. Value of confidence determines how frequently items in B appear in transactions that contain A [2]. These metrics are defined as [20]

$$\text{Support, S of } (A \rightarrow B) \;=\; \frac{\sigma(A \cup B)}{N} \tag{1}$$

$$\text{Confidence C of } (A \rightarrow B) \;=\; \frac{\sigma(A \cup B)}{\sigma(A)} \tag{2}$$

In mining, association rules are generated that satisfy user specified minimum support and minimum confidence thresholds. Through experimentation minimum support and confidence threshold values are set. To improve performance of association rule mining algorithms, support and confidence requirements should be decoupled [11]. To overcome this, Apriori algorithm [2] divides the problem into two tasks as: frequent item set generation and from the generated frequent item set, extraction of all the high-confidence rules. So, in the proposed system, Apriori algorithm is used for mining frequent item sets for generating strong association rules.

3. PROPOSED SYSTEM BLOCK DIAGRAM

Figure 1 shows the block diagram of the proposed method. Here, the novelty is in optimization of association rules using graph theory and Genetic algorithm with adaptive crossover and mutation rates using multi objective fitness function to give non-redundant, highly correlated and strong association rules.
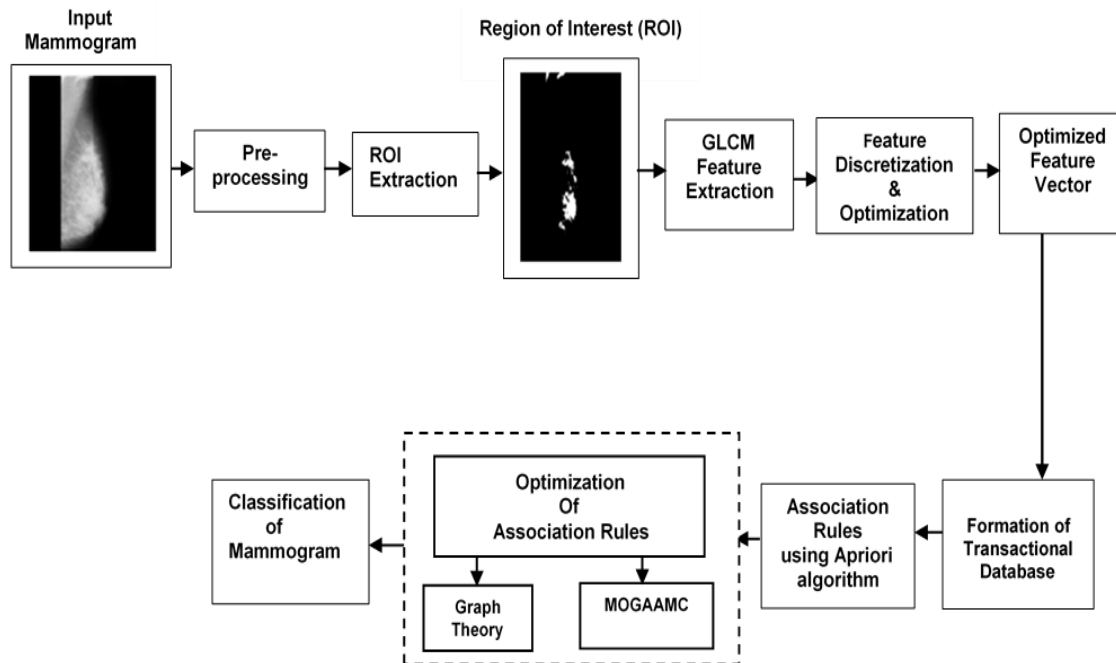
FIGURE 1: Block diagram of proposed image mining method.

3.1 Image Pre-processing and ROI detection

Standard mammography medical image dataset (MIAS/DDSM) is used for experimental analysis. Thresholding and Contrast Limited Adaptive Histogram equalization (CLAHE) are performed on the pre-processed mammogram for the removal of radiopaque artifacts. Finally, the images are segmented to find the region of interest (ROI).

3.2 Feature Extraction and Feature Vector Generation

Grey Level Co-occurrence Matrix (GLCM) is used to extract texture features from the segmented mammogram images. For every input Mammogram, GLCM matrix is generated. GLCM matrices [3] are calculated for the directions of 0^0, 45^0, 90^0 and 135^0. Every input image is expressed by its unique feature vector. Eight features namely correlation, entropy, energy, variance, homogeneity, 3rd Moment, invariance and sum of variance are calculated at different positions in continuation producing a feature vector of 32 elements to represent each image.

3.3 Feature Discretization and Selection

The feature vector and the labels of the input mammogram images are given as input to Optimum Feature Decomposition (OFD) algorithm. It is based on the PRESAGE algorithm reported in [4]. In this algorithm, discretization of continuous values of features and selection of feature is performed. Number of feature intervals and feature selection is based on the inconsistency in the feature values. [3]. The inconsistent and irrelevant features are removed and most discriminative features are selected to form the feature vector for each image. Optimum Feature Decomposition (OFD) algorithm discretizes continuous values of feature into different feature intervals with cut points. It generates   minimum number of feature value intervals with minimum inconsistency in the feature value.   Most discriminating features are automatically selected from the original feature vector which satisfies the following three criteria:
- The number of cuts (limits of an interval of values) in the feature should be minimum.
- The count of minimum cuts should be common in most of the features
- Selected optimized features should not affect the accuracy of the associative classifier.

Algorithm 1: Optimum Feature Decomposition Algorithm (OFD)

Input: A: Feature Vector with class, a minimum occurrence threshold, and minimum occupancy threshold   of major class in a given interval.

Output: Optimized Feature Decomposed Vector (OFDV)

- Sort the feature vector A in a non-decreasing order.
- Scan feature vector A. If the two consecutive images have different classes, insert a cut between them at the start and the end.
- If the number of images between two cuts is less than the minimum occurrence threshold, remove the higher cut.
- If two consecutive cuts have the same majority class, and if it occupies space more than the minimum occupancy threshold of major class in given interval, merge them.
- Save the remaining cuts.
- Order features as per the number of cuts.
- Select the features that satisfy the above three criteria.
- Save the selected features discretized in an OFDV.

3.4 Transaction Database Formation
Labels of training images i.e. benign or malignant and optimized feature vectors are used to build the transaction database.

3.5 Association Rule Mining
Transaction record of all input images is given as an input to Apriori algorithm, for generating association rules with user defined minimum support and minimum confidence. Since all the rules generated by association rule mining are not significant and important, we proposed optimization technique to reduce number of rules.

4. PROPOSED TECHNIQUES: GRAPH THEROY BASED APPROACH
Proposed optimization technique uses graph theory based an objective function to find the optimized rules generated from the training data set using Apriori algorithms. The objective function is to develop a set of rules when plotted as a graph leads to a simple graph with a simple walk. This graph essentially generates a set of rules which are non-redundant and optimized. The following algorithm presents the process of optimizing the rules,

ALGORITHM 2: Optimized Association Rule using Graph Theory (OARGT)

Input: A set of rules R of the form R_i: {R_1, R_2, ..R_n, S_1, S_2,…S_n, C_1,C_2,…..C_n} where
R_i : training data set rules .
S_l = min sup: user specified support of the rule.
C_i = min conf : user specified confidence of the rule.

Output: A set of optimized rules without redundancy.

Method:
V ←{set of vertices generated};
E←{set of Edges/Loops generated};
W_{vi} ←{weights of vertices};
W_{ei} ←{weights of edges}; Tg←{empirical threshold};
C_{nt}←{counter}; W_t←{weight allocated to vertex and edge}
D ←{difference between the set of rules}

for each rule R_i in R
          do {
             Di← R_i − R $_{(i+1)}$
                   }

```
            for each Rᵢ in  Dᵢ do {
                        if  (Dᵢ > Tg) then  count (i) ←1
                                else   count (i) ← 0
                         }
for each count  (i)  do{
            if   count (i) == count (i+1) and count ( i+1) ==1
                        then   W_vi ← W (t+1) and W_ei ←W (t+1)
                                else
                                     W_vi←W_t and W_ei←0
                                 }
Let i match the number of rules in R
for each i do {
            if  W vi == W (vi+1) and W_ei == 1 then C nt← C (nt+1) and    W_vi←C nt
                                 }
                for each Rᵢ do {
                        W_ei← W_ei / C nt
                                 }

            D_d← maximum of W_ei

            for each Rᵢ do {
                        if W_ei = D_d then delete rule Rᵢ from set R.
                                 }
```

The vertex V_i is generated for the first rule from the generated rule set R and a weight W_{vi} is allocated to it. The rule R_i and rule $R_{(i+1)}$ are compared by using bijection. The empirical threshold Tg is obtained, and the first rule R_i and the second rule $R_{(i+1)}$ are compared. If the number of matched values does not exceed the set threshold Tg, first rule R_i and the second rule $R_{(i+1)}$ are same and hence a loop e_i with weight W_{ei} is generated around the vertex V_i and the weight of the vertex V_i is incremented by 1. The next rule in the rule set R i.e rule $R_{(i+2)}$ is considered and matched with the rule $R_{(i+1)}$. If the matched values do not exceed the threshold Tg, it means that again the rule $R_{(i+2)}$ and rule $R_{(i+1)}$ are similar and hence a loop with weight i+2 is created around vertex V_i and the weight of the vertex V_i is incremented to i+2. Thus, the weight of the vertex is equivalent to the number of same rules and hence helps to detect redundancy.

If the matched values of the rule R_i and $R_{(i+1)}$ does exceed the threshold Tg, it implies both the rules are different and hence a new vertex V_{i+1} with weight i+1 is created and an edge is generated from V_i to V_{i+1}. This edge represents the next non-redundant rule that should be present in the optimized rule set. The rule $R_{(i+1)}$ are now matched with the rule $R_{(i+2)}$ and depending on the similarity or dissimilarity between the two rules a vertex is generated with appropriate weights. This process is repeated for all the rules in the generated rule set R. Thus, a graph G (V, E) with vertices V that represent non-redundant rules, edges E that represent the order of the non-redundant rules in the optimized rule set and loops that represent redundant rules is generated.

The next step is to remove redundant rules shown by loops in the graph G and check for the objective function. At the beginning of every iteration, generated graph is evaluated for objective function to check if the graph is simple with a simple walk. If the graph G contains loops then, the vertex V_i is considered and the ranking is assigned to every loop around Vi as per the following formula given below:

$$Rank\_E = \frac{Weight\ of\ the\ loop}{Weight\ of\ the\ vertex} \qquad\qquad (3)$$

This procedure is repeated for all vertices in graph G which assigns ranking to loops. In the next step, vertex V_i is considered again and the loop with the highest rank is deleted which leads to deletion of redundant rule or same rule from the rule set R. The weight of the vertex V_i is decremented by 1. Rankings are reassigned to loops. This procedure is repeated for every vertex

in graph G. In the next iteration, graph is again evaluated for the objective function. The proposed technique optimizes the rules by eliminating redundant rules and restoring the essential rules. The algorithm converges when the objective function is achieved i.e the simple graph with simple walk is obtained.

5. MULTI OBECTIVE GENETIC ALGORITHM WITH ADAPTIVE CROSSOVER AND MUTATION

5. 1 Genetic Algorithm

Association rule mining problem is to find all rules in data with support and confidence values greater than user specified minimum support (min.sup) and minimum confidence (min.con) threshold value. Transactional database and empirically decided minimum support and minimum confidence values are given to Apriri algorithm to generate rules. Information interpreted varies according to the rule. Rules having high value of support and confidence give conflicting or redundant information, making rule uninteresting. Thus, the confidence and support measures are insufficient for filtering out uninteresting association rules. To overcome this limitation, different correlation measures are used for augmenting the support and confidence framework for association rules [11]. This leads to correlation rules of the form,

  A→B [support, confidence, correlation measure]   (4)

In the proposed algorithm, for each rule, support, confidence and correlation measures such as interestingness, predictiveness, comprehensibility and lift are calculated. Since all generated rules are not significant and important, we are optimizing the rules by using genetic algorithm.

Genetic algorithms are important for discovering association as they work with global search to discover the set of items frequency with less complexity. The whole set of rules generated are considered 'population' in genetic algorithm. The initial population is produced using a random number generator. In our case, it is set of rules generated using Apriori algorithms. In the present work, we tried to solve the association rule mining problem with a Pareto based genetic algorithms. The first task is to represent the possible rules as chromosomes. For which a Michigan approach is used for encoding /decoding [11]. Here each chromosome represents a separate rule. In this case, we have to encode the antecedent and consequent parts separately. The next step is to find a suitable scheme for encoding /decoding the rules to /from binary chromosomes. Binary encoding is used in this approach. Binary value "1" represents presence of items and "0" represents absence of items in an item set. Chromosome length is fixed and it is equal to the number of distinct items (n) which is obtained from the transactions database. Given an item set of length K, all the gens in a chromosome are encoded as 0.

Genetic algorithms consist of following steps.
- Generate random population of n chromosomes.
- Calculate the fitness function f(x) of each chromosome x in the population
- Generate a new population by repeating following steps until the new population is completed

(a) Natural Selection: Fitness function is calculated for each rule. Based on the fitness function, some rules will be naturally selected. According to the fitness function, two parent chromosomes having better fitness will be selected.

(b) Crossover: Crossover is a significant feature of GA. Crossover obtains two individuals called parents and constructs two new individuals called offspring by swapping pairs of parents with a crossover probability. For crossover, select a pair of rules at random from those naturally selected, it selects a point along the length of the rule and swap the attributes of the two rules after that point. It checks whether the fitness function of this newly formed rule is above a predefined threshold. If it is, select the rule otherwise discard.

(c) Mutation: Mutation is used to maintain diversity from one generation of population to the next generation. Mutation changes one or more gene values in a chromosome from its previous generations. It mutates new offspring at each position in chromosome with the mutation

probability. For mutation, it selects a rule at random from those naturally selected and selects a point along the length of the rule and modifies the attribute at that point. Check whether the fitness function of this newly formed rule is above a predefined threshold. If it is, select the rule otherwise discard. Normally mutation probability is set to low. If it is too high, search will become a primitive random search. The adaptive mutation gives better performance than a fixed mutation rate [17].

(d) Accepting: Place new offspring in a new population
Use the newly generated population for a further processing of algorithm.
If the end condition is satisfied, stop further processing of the algorithm, and return the best solution in current population.

(**e) Termination Criteria**
The termination criteria are the criteria by which the GA decides whether to continue searching or to stop the search. The possible termination conditions are listed below [12]
- Fixed number of generations reached.
- The solution's fitness with highest ranking at a fixed number of generations
- Manually inspecting the solutions
- Combination of the above.

5.2. Multi Objective Fitness Function
Authors use Multi objective Genetic algorithm to optimize association rules generated using Apriori algorithm using interestingness correlation measures as fitness function rather than support and confidence to obtain strong association rules. It is interactive and iterative process, with many decisions being made by user. Existing algorithms try to measure the quality of generated rule by considering only one evaluation criteria. This criterion evaluates rule depending on number of occurrence of rule in entire database. The generated rule may have many attributes involved in the rule there by making it difficult to understand. Since more importance is given to those rules, satisfying number of records.
Here we considered that discovered knowledge must have four general properties [10, 15, and 18] such as Comprehensibility, Predictive Accuracy, Interestingness and lift. They are defined as below:

$$\text{Comprehensibility} = \text{CR} = \frac{\log|1+|B||}{\log|1+|A\cup B||} \tag{5}$$

Here, {B} and {A U B} are the number of attributes involved in the consequent part and the total rule respectively.

Predictiveness: It is measured by using following expression

$$P_{acc} = \frac{|A\&B|}{|B|} \tag{6}$$

Where |A&B| = No. of record satisfying both A and B

Interestingness: Interestingness of a rule is used to measure importance of the rule to the user. Interestingness can be defined by using the following expression [10, 13].
Interestingness of A→B,

$$\text{Intt} = \frac{SUP(A\cup B)}{SUP(A)} \times \frac{SUP(A\cup B)}{SUP(B)} (1 - \frac{SUP(A\cup B)}{\sigma(N)}) \tag{7}$$

Where σ (N) is the total number of records in the database.

$$Lift = \frac{SUP(A \cup B)}{SUP(A)\, SUP(B)} \tag{8}$$

Overall multi objective fitness function is calculated by considering all above parameters, and taking their arithmetic means as

$$F = \frac{w_1 P_{acc} + w_2 C(R) + w_3\ Intt + w_4\ Lift}{w_1 + w_2 + w_3 + w_4} \tag{9}$$

Where w_1, w_2, w_3 and w_4 are user defined weights. In the proposed method, the rules generated are tested for the fitness function defined by the equation (9).

5.3 Proposed MOGAACM Algorithm

There are some difficulties to use the standard multi-objective GAs for association rule mining problems. In case of rule mining problems, we need to store a set of better rules found from the database. If we follow the standard genetic operations, then the final population may not contain some rules that are better and generated at some intermediate generations. For this task, a separate population is used [11]. In these populations, no genetic operations are performed. It will simply contain only the non-dominated chromosomes of the previous generations. The user can fix the size of populations. At the end of first generation, it will simply contain the non-dominated chromosomes of first generations. After the next generations, it will contain non-dominated among the current populations plus non-dominated solutions till the previous generations. Genetic algorithm itself provides good solution. But every iteration considers the parameters of the first generation only. Also, the number of new offspring for high fitness rules or individuals is always high and for low fitness it is low. So, there is a need for adaptive approach which resets the parameters after every iteration [19] and treats each class of individuals separately. In the CAD model, ranking of rules is done based on fitness function. The crossover and mutation rate are low for higher ranked rules and vice versa. Also, a parameter α is defined which is updated after each iteration. It defines the minimum threshold of fitness function needed for the new progeny to be selected. It depends on the new set of rules generated after every iteration, so the whole process becomes adaptive as α adapts itself according to the new set of rules generated. The entire process is explained below.

ALGORITHM 3: MOGAACM Algorithm

- Initialize the population with the ARM rules.
- Define the fitness function of each rule.
- Perform the natural selection of rules using Roulette Wheel Selection.
- Rank the selected rules according to the fitness function i.e. the rules with the highest fitness function will be assigned high rank & vice versa.
- Set the minimum fitness function of these selected rules as the threshold say α, which is adapting itself after every iteration, according to the new set of rules generated by GA at the output after each iteration.
- Perform crossover between randomly selected pairs of rules at specified number of times. The crossover rate depends on the rank of rules i.e. the high rank rules have low crossover rate & vice versa. Check the fitness function of each & every new rule & if it above the threshold α, select the rules, otherwise we discard.
- Perform mutation on randomly selected rules a specified number of times. The mutation rate depends on the rank of rules i.e. the high rank rules has low mutation rate & vice versa. Check the fitness function of each & every new rule & if it above the threshold α, select the rules, otherwise we discard.
- Combine the selected, cross overed & mutated rules into Newset_rules. This completes iteration.
- Input Newset_rules to Step 1. Carry out such iterations number of times as specified by the user.

A transactional database of the testing images is created and compared with obtained rules from ARM. Following algorithm is used to classify the testing image into appropriate class.

6. CLASS IDENTIFICATION USING STRENGTH OF CLASSIFICATION ALGORITHMS

Strength of classification is an essential component in the ARM classification required at the final classification stage.   Since the proposed ARM produces sensitive rules, the antecedent part always defines the final class; thereby segregation of the rules based on the antecedent part needs to be done. Thus, finally a set of rules, each belonging to different class is obtained. Let T(i) is denoted as the total number of rules which contain class 'i' as the antecedent part. The testing transaction database is retrieved by considering one transaction at a time. The numbers of rules that are fully satisfied by the given transaction are calculated.

A rule is said to be 'fully-satisfied' by a transaction, if the transaction contains all the antecedent components of the rule in it. It is denoted by F(i, j), considering j^{th} transaction and i^{th} class. The numbers of rules that are partially satisfied by the given transaction are also calculated. A rule is said to be 'partially-satisfied' by a transaction, if the transaction contains at least, two antecedents components of the rule in it. It is denoted by P (i, j), considering j^{th} transaction and i^{th} class. The strength of classification for each class for each transaction is calculated as:

$$SC(i,j) = \frac{F(i,j) + w*P(i,j)}{T(i)} \tag{10}$$

Where, SC(i, j) : Strength of classification for j^{th} transaction belonging to class i.
F(i, j) : Number of fully satisfied rules for j^{th} transaction belonging to class i.
P(i, j) : Number of partially satisfied rules for j^{th} transaction   belonging to class i
w: Weight assigned for the rules which are partially satisfied.
T(i): Total number of rules belonging to class i.

Class identification: In this transaction is loaded, SC for each class is calculated and compared. The highest value of SC is selected and the class to which it belongs is the class which must be assigned to the test transaction.

> For i=1: Class is Benign
> For i=2: Class is Malignant
> If SC (i = 1; j) ≥ SC (i = 2; j),
> Transaction is classified as Benign.
> Else
> Transaction is classified as malignant.

7. EXPERIMENTAL RESULTS

7.1 Graph Theory based Objective Function Technique

Mammography Image Analysis Society (MIAS) Database is used for experimentation. It consists of 322 images, corresponding to two classes: normal (208) images, abnormal containing benign (63) and malignant (51). All these images include the locations of any abnormalities present. This database is selected, as it is freely available and standard database and able to compare proposed method with other published work from the database. Number of images used for testing and training are represented in Table1.

Training images		Testing images	
Benign	Malignant	Benign	Malignant
50	40	13	11

TABLE 1: Images used for Experiment.
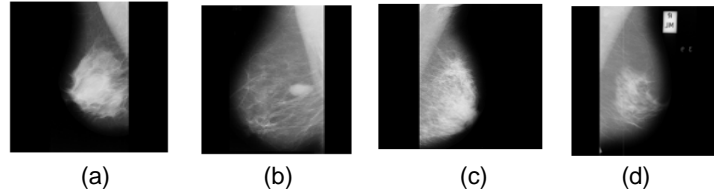
(a) (b) (c) (d)

FIGURE 2: Sample original mammogram Images from MIAC database.
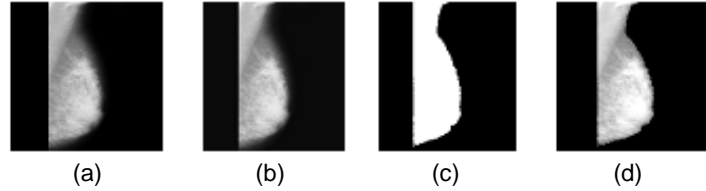

(a) (b) (c) (d)

FIGURE 3: Preprocessing steps: (a) original image (b) filtered image (c) thresholded image (d) image post contrast enhancement.

Image No	Image Features from 1 to 32					
	1	2	……..	22	..	31	32
1	0.995104	0.99263		15.54366	.	42.98692	42.79023
2	0.995871	0.993921		16.2128	.	45.32159	45.13118
3	0.992838	0.989539		27.87123		81.21449	80.84955
4	0.993488	0.990719		25.28826		73.21357	72.84488
5	0.993907	0.991231		23.90241		71.97366	71.71099
6	0.993289	0.989421		27.06293		78.58713	78.21574
7	0.993818	0.992166		26.67008		77.54006	77.21209
8	0.994269	0.989931		29.31752		87.36543	86.98926
.							
.	0.994454	0.991078		16.75225		51.11041	50.88733
.	0.992504	0.989593		16.94954	.	51.00482	50.73681
.	0.991824	0.985348		14.43449	.	42.11944	41.87267
90	0.990957	0.985042		13.49663	..	39.17426	38.95321

TABLE 2: GLCM Feature Vector.

Image Label	Optimized feature value interval label (1 to 27)									
	1	2	3	4	5	…….	24	25	26	27
11111	10	21	32	43	56	………….	321	335	346	356
11111	10	22	34	46	56	.	323	335	346	358
11111	4	16	27	38	50	.	316	329	341	352
11111	5	18	29	41	51		316	329	341	352
11111	7	19	29	41	53		317	329	341	352
11111	5	16	27	40	51		313	326	339	351
11111	6	20	32	44	52		319	332	345	355
11111	8	17	28	40	54		314	325	338	349
…..										
22222.	9	21	33	45	55		320	334	345	357
22222.	9	18	30	42	55	.	320	332	344	354
22222.	3	16	27	39	49	.	318	330	342	353
22222	2	22	23	35	48	………….	318	326	338	348

TABLE 3: Transactional Database.

	Rules	Description
R$_1$	21,36,70,9 → 11111	Image having optimized feature value interval label as 21, 36, 70 98 tends to a benign image
R$_2$	30,76,88,320→33333	Image having optimized feature value interval label as 30,76,88,320 tends to be malignant image

TABLE 4: some examples of association rules mined.

Grey Level Co-Occurrence Matrix (GLCM) method is used for automatic extraction of textual feature from the segmented mammogram ROI images. These features are organized into feature vectors. Each image is represented by feature vector. Table 2 shows feature vector where in the first value in each row is image number and second onward is its feature values. These features are discretized and selected. In this step, feature selection and discretization of selected features is done. The most discriminative features are selected to form optimized feature vector. Total 32 features are obtained from feature extraction step for each input image. From these 32 features, 27 most discriminative features are selected. Feature vector and the labels of the input mammogram images form a transactional database. Table 3 shows transactional database where first column indicate image label second column onwards gives the values of 27 optimized feature value interval labels. The transaction database for training images is given as input to Apriori algorithms for association rule mining. Table 4 shows some examples of association rules mined.

Parameters:
mini sup threshold = 3 %,
mini conf = 100%
Generated rules = 3251 for benign class
Generated rules = 922 for malignant class

For testing images, features are extracted, optimized and discretized into intervals and given as input to classifier. For classification, association rules generated for training images and feature vector of test image are used. Classification algorithm stores keywords belonging to consequent part of association rules in database. Feature vector of the test image is compared with antecedent part of all association rules. Match is found if value of test image feature lies in range of feature value of benign or malignant mass. On class by class basis, scores are incremented, when match to respective class is found. Test image is classified in to the class having highest matching score. Classification step gives the class of test image i.e. benign or malignant.

 All these rules are not significant and redundant. So, graph theory based objective function is used to extract of strong and non-redundant rules. The difference in magnitude of cut points of discretize feature is defined as 'threshold' and is varied between the intervals of 5 to 90 and number of non-redundant rules are calculated. It is observed that at Threshold = 5, number of non-redundant   rules are generated = 2958 and 829 for benign and malignant class respectively. With these rules, same, we obtained the same accuracy with the original rules. So, using graph theory technique there is 9-10% reduction in number of rules. Table 5 shows classification accuracy.

Technique	Apriori		Graph Theory	
	Rules	Accuracy	Rules	Accuracy
Benign	3251	92.307	2958	92.307
Malignant	922	90.909	829	90.909

TABLE 5: Classification Summary.

Category		Predicted Class	
		Benign	Malignant
Actual Class	Benign	T. P = 12	F.P = 01
	Malignant	F.N = 01	T.N = 10

TABLE 6: Confusion Matrix for Graph Theory.

A confusion matrix contains information about actual and predicted classifications done by a classification system. Table 6 shows confusion matrix which represents the Contingency table, where True Positive (TP) means normal cases classified correctly. False Positive (FP) means abnormal cases classified as normal. True Negative (TN) means abnormal cases classified correctly. False Negative (FN) means normal cases classified as abnormal by classifier. We also estimated accuracy 91.60 % sensitivity 92.3 %, specificity 90.99 %, and precision 92.30 % and recall 92.30 %

7.2 Experimentation Using MOGA with Adaptive Mutation and Crossover
Simulation in MATLAB environment is carried out to validate the proposed algorithm. We selected mammographic images from two databases, Mammographic Image Analysis Society (MIAS) database [2] and Digital Database for Screening Mammogram (DDSM) [3]. Table 7 shows number of training and test images used for MIAS and DDSM dataset.

Database	Training Images		Testing Images	
	Benign	Malignant	Benign	Malignant
MIAS	50	40	13	11
DDSM	78	78	40	40

TABLE 7: Training and Testing Images.

Image No	Image Feature Vector (1 to 32)						
	1	2	3	..	30	31	32
1	0.9954	0.9929	0.9912	..	0.9983	0.9981	0.9981
2	0.9963	0.9945	0.9934	..	0.9984	0.9983	0.9983
3	0.9967	0.9944	0.9927	…	0.9986	0.9985	0.9985
4	0.9958	0.9937	0.9924	..	.	.	.
5	0.9966	0.9944	0.9930	..	.	.	.
.	.	.	.	..	.	.	.
.	.	.	.	..	.	.	.
.		.	.	..			.
154	0.99589	0.9946	0.9941		0.9993	0.9992	0.9992
155	0.9949	0.9938	0.9933		0.99924	0.9991	0.9990
156	0.9958	0.9940	0.9926		0.9992	0.9991	0.9991

TABLE 8: Feature vector for DDSM.

Image No	Optimized feature value interval label (1 to 27)									
	1	2	3	4	5	6	..	25	26	27
1	16	75	107	43	56	64	..	907	935	959
2	25	89	122	46	56	66	..	915	943	969
3	26	83	114	38	50	45	..	914	939	962
..	..	.	.	.	.	.	.	.	.	.
..	..	.	.	.	.	.	.	.	.	.
153	9	33	33	45	55	61	..	334	345	357
154	21	113	30	42	55	62	..	926	956	978
155	11	117	27	39	49	57	..	922	954	974
156	20	120	23	35	48	67	..	923	952	975

TABLE 9: Feature discretization and selection for DDSM.

For DDSM data set, Table 7, Table 8 and Table 9, represents feature Vector, feature dissertation and selection and Transactional database respectively. For DDSM dataset, Apriori algorithms parameters are as mentioned below:

min. sup = 7
min. con= 90,
Generated rules for benign class = 2019 and Generated rules for malignant class = 2996

Image Label	Optimized feature value interval label (1 to 27)											
	1	2	3	4	5	6	..	..	24	25	26	27
11111	16	75	107	43	56	65	..	..	890	907	935	959
11111	25	89	122	46	56	64	..	..	896	915	943	969
11111	26	83	114	38	50	50	..	..	900	914	939	962
..	..	.	.	.	.		.	.		.	.	.
..	..	.	.	.	.		.	.		.	.	.
22222	9	33	33	45	55	67	..	.	278	334	345	357
22222	21	113	30	42	55	76	..	.	867	926	956	978
22222	11	117	27	39	49	59	..	.	879	922	954	974
22222	20	120	23	35	48	60	..	..	899	923	952	975

TABLE 10: Transactional Database for DDSM dataset.

Data Set	Algorithm	Parameters	No. of rule generated		Accuracy		
			Benign	Malignant	Benign	Malignant	Avg
MIAS	Apriori	Min. sup = 7, Min. conf = 90	3251	922	92.307	90.90	91.666
DDSM	Apriori	Min. sup = 3, Min. conf = 90	2019	2996	95	95	95
MIAS	Graph Theory	Min. sup = 3, Min. conf = 90, Threshold Tg =0.01 Difference between cut point of optimized feature vector:5	2958	829	92.307	90.90	91.666
MIAS	MOGAACM	Initial Population =3251 Initial Population =922 Crossover Probability = 0.75 Mutation Probability = 0.005 Threshold α = 0.001 crossover = Single Point	355	520	100	90.90	95.45
DDSM	MOGAACM	Initial Population = 2019 Initial population= 2996 Crossover Probability = 0.75 Mutation Probability = 0.005 Threshold α = 0.001 Crossover = Single Point	401	611	92.5	92.5	92.5

TABLE 11: Comparison of Algorithm: Apriori, Graph theory and MOGAAMC.

MOAGAAMC algorithm is used to extract strong and optimized rules. Experimentation is carried up to 6th iteration. It is observed that for iterations 4, 5 and 6, association rules are nearly same. So we stopped at 6th iterations. At 6th iteration, 401 and 611 strong rules are obtained for benign and malignant class respectively. A confusion matrix contains information about actual and predicted classifications done by a classification system. The same algorithm is tested on MIAS database.

Table 11 shows results of Apriori, Graph Theory and MOGAAMC on DDSM and MIAS database. Table 12 shows confusion matrix for both database. We use sensitivity, specificity, precision and recall for estimating accuracy of the mammogram classification. Table 13 shows classification accuracy comparison of proposed algorithms with algorithms reported in literature.

MIAS				DDSM			
Category		Predicted Class		Category		Predicted Class	
		Malignant	Benign			Malignant	Benign
Actual Class	Malignant	TP = 10	FN = 1	Actual Class	Malignant	TP = 37	FN = 3
	Benign	FP = 00	TN = 13		Benign	FP = 3	TN = 37

TABLE 12: Confusion Matrix for MIAS and DDSM Database.

Investigator	Features	Classifier	Database	Accuracy
L.M.Bruse [21] [1999]	DWT Shape Features	LDA	60 Images	83 and 80 parent and live
Maria-Luiza[2001] [07]	Shape and GLCM	BPNN	MIAS	92
Sumeet Dua et al [22] [2009]	Texture	Weighted Associative classification	MIAS	89
Wang D et al.[2009] [23]	GLCM, Gabor	Structured SVM	DDSM	91.4
W. Borges Sampaio, et al. [2011] [24]	Shape and Texture	SVM	DDSM	80
Deepa Deshpande [25][2013]	Texture features	Association Rule mining	MIAS	100 Benign 84 malignant
Zhili Chen [26] [2015]	Shape and MRF Density	Density Map	MIAS and DDSM	MIAS: 76.01 DDSM :81.22
Gisele Helena [27] [2015]	Counter Shape and density features	Fuzzy Rule	DDSM	83.34
Yanfeng et al. [2015] [28]	Texton with multiple sub sampling analysis	KNN	DDSM	85.96
Kanchanamani et al. [2016] [29]	Shearlet transform and statistical features	SVM	MIAS	92.5
Wang J. et al. [2016] [30]	Statistical and textual features	Deep learning	Clinical Database	87.3
Proposed Approach	**GLCM**	**Proposed ARM With MOGAACM**	MIAS DDSM	**MIAS: 95.45 DDSM:92.5**

TABLE 13: Comparison of proposed algorithms performance with algorithms reported in literature in terms of accuracy.

8. CONCLUSION

The authors presented image mining technique based on association rule mining algorithm for extracting image frequent patterns and hidden information in Mammography image database. Two approaches are proposed to optimize the association rules. The first, graph theory based on objective function such that graph generated by the optimized rules is simple graph with simple walk. The second is MOGAAMC. The generated rules are interesting, easy to understand and maximize the fitness function based on four measures namely Comprehensibility, Predictive Accuracy, Interestingness and lift. The proposed approach introduces reinitialization process and adaptive crossover and mutation to perform evolutionary learning and to improve diversity of obtained rules. The obtained rules are strong, showing a strong relationship among the feature sets and reducing the problem of support-confidence framework. Experimental results show that, MOGAACM generates more effective and strong association rules compared with objective function using graph theory and achieves 89.08% and 43.60 % reduction in association rules for benign and malignant class respectively for MIAS database and 80.13 % and 79.60 % reduction in association rules for benign and malignant class respectively for DDSM medical image database. The Proposed CISCA algorithm efficiently classifies the input mammogram with classification accuracy 91.66 % for MIAS database using graph theory, 95.45 % and 92.5 % for MIAS and DDSM respectively using MOGAACM.

9. ACKNOWLEDGEMENT

10. REFERENCES

1. J. Zhang, W. Hsu and M. L. Lee. "Image Mining: Trends and Developments." Journal of Intelligent Information Systems, 19:1, pp. 7-23, 2002.

2. R. Agrawal et al., "Mining association rules between sets of items in large databases", in proceedings of the   ACM SIGMOD ICMD, Washington DC,1993, pp 207-216

3.   R. M. Haralick et al.," Textural features for image classification", IEEE Trans Syst. Man. Cybern., Vol. SMC-3.610-621, 1973.

4. M. X. Ribeiro, J. M. Agni,T. Caetano , and P. M. Azevedo-Marques, "An Association Rule-Based method to support medical images diagnosis with efficiency.", IEEE Transactions on Multimedia, Vol.10, No. 2, pp. 277-285, February 2008.

5. A. K. Mohanty, S. K. Swain, P. K. Champati, S. K. Lenka, "Image Mining for Mammogram Classification by Association Rule Using Statistical and GLCM features", International Journal of Computer Science Issue, 2011.

6. M. L. Antonie, R. Osmar, Zaiane, C. Alexandru, "Mammography classification by an association rule-based classifier", in Proceedings of Third International Workshop on Multimedia Data Mining, 2002, pp. 62-69.

7. M. L. Antonie, R. Osmar, Zaiane, C. Alexandru, "Application of Data Mining Techniques for Medical Image Classification", in Proceedings of Second International Workshop on Multimedia Data Mining (MDM/KDD'2001) in conjunction with ACM SIGKDD conference, San Francisco, USA, Aug 26, 2001; pp. 94-101.

8. A. Dhawan et al.," Analysis of mammographic micro calcification using gory scales level image structure features, IEEE Transactions on medical imaging, 15(3), 246-259, 1996

9. B. Pratibha, B. Sadhashivam, "Brest Tissue characterization using variant of nearest neighbor classifier in multi texture domain", IE(1), J91, pp 7-13, 2010.

10. A. Ghosh, B. Nath," Multi-objective rule mining using genetic algorithms", Information Sciences, Elsevier, Vol.163, pp.123–133, 2004.

11. X. Yan, C. Zhang and S. Zhang," Genetic Algorithms based strategy for identifying association rules without specifying actual minimum support", Expert System with Applications, Science Direct, 36(2009) page 3066-76.

12. S. Kannimuthu, K. Premalatha, "Discovery of high utility item sets using genetic algorithm with ranked mutation. Appl. Artificial Intelligence. 28(4), 337–359 (2014)

13. M. U. Bandyopadhyay and S. Mukhopadhyay," A Survey of Multi objective Evolutionary for data mining", IEEE Transactions E Vol, Commutating, 18(1),4-19, 2014.

14. R. D. Martin, A. Fedz," A new Multi objective Evolutionary Algorithms for mining a reduced set of Interesting Positive and negative quantities Association rules", IEEE Transactions Evolutionary Commutating 18(1),54-69, (2014)

15. U. Maulik, S. Bandyopadhyay and A. Mukhopadhyay, "Multiobjective Genetic Algorithms for Clustering Applications in Data mining and Bioinformatics", Springer, Heidelberg-Berlin, ISBN 978-3-642-16614-.

16. P. Shenoy, K. Srinivas, K. Venugopal and L. Patnaik, "Dynamic Association rule mining using Genetic Algorithms ", Intel data Analysis, 9(5), 439-453 (2005).

17. K. Premalatha and A. M. Natarajan," Genetic Algorithms for document Clustering with Simultaneous and ranked Mutation", Journal of Modern Applied Science, 3(2), 2009, Page 75-82.

18. S. Dehuri, A. Ghosh, R. Mall," Genetic algorithm for Multi-criterion   classification and clustering in Data mining", International Journal of Computing and Information Science, Vol 4, No. 3, Pages 143-154.

19. Mir Md. J. Kabir, X. Shuxiang, B. Ho Kang, and Zongyuan Zhao," Discovery of Interesting Association Rules Using Genetic Algorithm with Adaptive Mutation", ICONIP 2015, part II, LNCS, pp 96-105, 2015.

20. J. Han, M. Kamber, J. Pei, "Data Mining Concepts and Techniques", Morgan Kaufmann publishers, 2012.

21. L.M. Bruce, R.R. Adhami, classifying mammographic mass shapes using the wavelet transform modulus-maxima method, IEEE Trans. Med. Imaging 18 (1999) 1170-1177

22. D. Sumeet, S. Harpreet, H.W. Thompson, Associative classification of mammograms using weighted rules, Expert Systems with Applications, ELSEVIER, (2009), 36, 9250–9259

23. D.Wang, Shi L, Ann Heng P, Automatic detection of breast cancers in mammograms using structured support vector machines, Neurocomputing2009; 72: 3296-3302.

24. W.B. Sampaio, Diniz EM, Silva AC, de Paiva AC, Gattass M." Detection of masses in mammogram images using CNN, geostatistical functions and SVM. 653-664 Article in Computers in Biology and Medicine 41(8):653-64 · June 2011

25. D. S. Deshpande, A. M. Rajurkar and R. M. Manthalkar, "Medical image analysis an attempt for mammogram classification using texture based association rule mining," 2013 Fourth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), Jodhpur, 2013, pp. 1-5.

26. C. Zhili, H. Strange, A. Oliver et al., "Topological Modeling and Classification of Mammographic Macrocalcification Clusters", IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, Vol. 62, No. 4, April 2015, pp. 1203-1214.

27. H. Gisele, M. Miranda, C. Joaquim ", Computer-aided diagnosis system based on fuzzy logic for breast cancer categorization", Computers in Biology and Medicine 64 (2015) 334-346, Elsevier.

28. L. Yanfeng et al.," Textron analysis for mass classification in mammograms", Pattern Recognition Letters, Volume 52, pp. 87-93

29. M. Kanchanamani, P. Perumal, "Performance evaluation and comparative analysis of various machine learning techniques for diagnosis of breast cancer", proceedings of biomedical research 2016.

30. J. Wang, Yang X, Cai H, Tan W, Jin C, Li L.," Discrimination of Breast Cancer with Macrocalcifications on Mammography by Deep Learning". *Scientific Reports*. 2016; 6:27327.