International Journal of Computational Intelligence and Informatics, Vol. 4: No. 1, April - June 2014

Mammogram Image Classification: Non-Shannon Entropy based Ant-Miner

R Roselin
Department of Computer Science, Sri Sarada College for Women (Autonomous), Salem, Tamil Nadu, India.

K Thangavel
Department of Computer Science Periyar University Salem, Tamil Nadu, India.

C Velayutham
Department of Computer Sciecne Aditanar College of Arts and Science Tiruchendur, Tamil Nadu, India.

***Abstract*- Mammography plays the central role in the early detection of breast cancer because it can show changes in the breast up to two years before a patient or physician can feel them. Mammogram is a radiograph of the breast tissue. Finding the breast tumor before they turn deadly is a challenge and that the medical technology so far failed master. A newly developed system should help radiologist in more accurate diagnosis. This paper applies data mining technique on mammogram image processing, more specifically it applies swarm intelligence based classification algorithm called Ant-Miner. The original Ant-Miner algorithm uses Shannon Entropy in its heuristic function. A novel idea of using Non-Shannon entropy measure in the heuristic function has been analyzed because of its non-extensiveness and a comparative analysis is made with the decision tree algorithm C4.5 in terms of accuracy, number of rules, True Positive Rate (TPR), and False Positive Rate (FPR) and it is reported that Tsallis entropy based Ant-Miner has been proposed for its performance in mammogram image classification.**

Keywords - Mammogram, Ant-Miner, Entropy, Shannon Entropy, Non-Shannon Entropy, Swarm Intelligence

I. INTRODUCTION

The effectiveness of mammogram in the detection of breast cancer is currently under investigation. Variety of algorithms has been developed over the years to process mammograms and to allow more accurate diagnosis by the radiologist. Most of the techniques in the literature concentrate on image processing techniques to enhance the mammogram and then go for classification through neural networks (Varela C. et al. 2007 ; Li H. et. al. 2008 and Fauci F. et al 2004) [1, 2]. More over Neural Networks require a number of parameters that are typically best determined empirically, such as the network topology or structure. Neural Networks have been criticized for their poor interpretability (Jiawei Han et. al. 2006)[3]. For this reasons an attempt is made to apply image mining techniques on mammograms. Image mining deals with extraction of implicit knowledge, image data relationships or patterns that are not explicitly stored in the image. In this paper swarm intelligence based classification algorithm called Ant-Miner has been applied and heuristic function is modified using Non-Shannon Entropy measures.

In colony of social insects such as ants, bees and wasps each insect perform its own tasks independently of each other but the task performed by these insects are related to each other in such a manner that a colony as a whole is capable of solving complex survival related problems through cooperation. This collective behaviour which emerges from these social insects has been called Swarm Intelligence (Bonabeau E. et al. and Dorigo M et al 1999) [4]. Ant Colony Optimization (ACO) is a branch of swarm intelligence interested in the behavior of natural ants. Real ants are capable of finding the shortest route between the food source and the nest without using any visual information. It uses the chemical substance called pheromone trails that help the successive ants to follow the shortest path between the food source and the nest (Dorigo M. et al. 1996) [5].

The design of ACO algorithm requires the appropriate representation of the problem. It requires the problem dependent heuristic function, probabilistic transition rule and pheromone updating function. Classification is the process of assigning the object in the data set into predefined classes. Mammogram taken from mini-Mammographic Image Analysis Society Database (MIAS) has been used for generating the knowledgebase (http://peipa.essex.ac.uk). It contains 322 mammograms with 208 normal cases, 63 benign and 51 malignant cases.

Data mining cannot be directly applied on the images. In order to apply data mining techniques on mammogram images, image needs to be converted to feature vector. Grey Level Co-occurrence Matrix (GLCM), a well-established, robust statistical tool for extracting second order texture information has been used and Haralick fourteen features are extracted from images.

Before extracting features a lot of pre-processing has to be done on the image. Mammogram image may contain artifact and pectoral muscles that may reduce the rate of accuracy in the classification model. Hence they need to be identified and removed before segmentation. Segmentation is the process of dividing the image into constituent part and extracting those of interest. Classification model is generated based on the features extracted from the segmented image by applying C4.5, Ant-Miner using Shannon Entropy measure as heuristic function and Ant-Miner using Non-Shannon entropy measures and the comparative analysis is done.

The rest of the paper is organized as follows: Section 2 briefly reviews the related work. Section 3 describes the original Ant-Miner algorithm with Shannon entropy measure as heuristic function. Section 4 explains the Non-Shannon entropy measures. Section 5 provides the experimental setup. Section 6 shows the experimental results, followed by Section 7 discusses the comparative analysis of both the Ant-Miner in terms of accuracy, number of rules, True Positive Rate (TPR) and False Positive Rate (FPR). Finally, Section 8 concludes our paper.

II. RELATED WORKS

Most of the existing methods in the literature uses neural network as the classifier (Varela C. et al. 2007 ; Li H. et. al. 2008 and Fauci F. et al 2004)[1, 2, 6]. Jiang, J. et al. (2007) [7] used genetic algorithm to classify the mammograms. Stylianos et al. (2011) [8], Jinchang Ren (2012) and Juan F. et al. (2012) [9, 10] used support vector mission for mammogram classification.

The first ACO based algorithm for classification rule discovery, called, Ant Miner (Parpinelli et al. 2002)[11]. The information gain (Entropy) has been used as the heuristic value of a term. After the antecedent part of a rule has been constructed, the consequent of the rule is assigned by a majority vote of the training samples covered by the rule. The constructed rule is then pruned to remove the irrelevant terms and to improve its accuracy.

The extensions of the Ant Miner algorithm were proposed by Liu et al. in Ant Miner2 (Liu et al. 2002)[12] and Ant Miner3 (Liu et al. 2004)[13]. Ant Miner2 uses density estimation as a heuristic function instead of information gain used in Ant Miner. They showed that this heuristic value does the same job as well as the complex one and hence Ant Miner2 is computationally less expensive than the original Ant Miner. Ant Miner3 uses a different pheromone update method with the pheromone of only those terms that occur in the rule and do not evaporate the pheromones of unused terms. In this way exploration is encouraged.

David Martens et al. (2007) [14] proposed a Max-Min ant system based algorithm (AntMiner+) that differs from the previously proposed Ant Miners in several aspects. Only the best ant is allowed to update the pheromone, the range of the pheromone trail is limited within an interval, class label of a rule is chosen prior to the construction of the rule and a different rule quality measure is used.

Other works on Ant-Miner include (Smaldon J. et al. 2006)[15] in which an algorithm for discovering unordered rule sets has been presented. PSO (Holden N. et al. 2007)[16] algorithm is used for continuous valued attributes and ACO for nominal valued features and these two algorithms are jointly used to construct rules. The issue of continuous attributes has also been dealt in (S. Swaminathan 2006 and Otero F. et al. 2008) [17, 18]. The proposed algorithm retains the basic structure of the previous Ant Miner algorithms. In this proposed method, heuristic function uses Tsallis entropy instead of Shannon Entropy measure. A comparative study is made in the field of Mammogram image processing.

III. THE ANT-MINER ALGORITHM

Ant Colony optimization is a branch of swarm intelligence inspired by the behaviour of natural ants. ACO algorithms are based on the aspect of the food foraging behaviour of ants. As ants move a certain amount of pheromone is dropped on the ground, marking the path with trail of this substance. The more ants follow the given trail, the more attractive this trial becomes to be followed by other ants. This process can be described as a loop of positive feedback, in which the probability that an ant chooses a path is proportional to the number of ants that have already passed by that path. This indirect form of information passing via the environment helps the ants to find the shortest path to the food source. If two paths between a food source and the ant nest are initially discovered by some ants, then the longer of the two paths will soon become unattractive to subsequent ants because the ants following it will take longer to go to the food source and return back and hence the pheromone concentration on that path will not increase as rapidly as on shortest path. This phenomenon has been modelled in the ACO algorithm.

An artificial ant constructs a solution to the problem by adding solution component one by one. When solution is constructed its quality is determined and the components of the solution are assigned pheromone concentration proportional to the quality. Subsequently other ants construct their solutions one by one and they are guided by the pheromone concentration in their search components to be added in their solutions. The pheromones with higher concentrations are thus identified as contributing to a good solution and repeatedly appear in the solutions. It is expected that after a while the ants converge on a good, if not the optimal solution.

Since its inception, ACO can be applied to solve many combinatorial optimization problems such as, quadratic assignment, job scheduling, subset problems, network routing, vehicle routing, load dispatch power systems, bioinformatics, and of course data mining. For the application of ACO to a problem involves the specification of

- An appropriate representation of the problem, which allows the ants to incrementally construct/modify solution through the use of a probabilistic transition rule, based on the amount of pheromone in the trail and on a local, problem-dependent heuristic.
- A method to enforce the construction of valid solutions, that is, solutions that is legal in the real-world situation corresponding to the problem definition.
- A problem-dependent heuristic function (h) that measures the quality of items that can be added to the current partial solution.

A rule for pheromone updating, which specifies how to modify the pheromone trail τ. A probabilistic transition rule based on the value of the heuristic function (h) and on the contents of the pheromone trail τ that is used to iteratively construct a solution.

Artificial ants have several characteristics similar to real ants, namely:

- Artificial ants have a probabilistic preference for paths with a larger amount of pheromone.
- Shorter paths tend to have larger rates of growth in their amount of pheromone.
- The ants use an indirect communication system based on the amount of pheromone deposited on each path.

A. *General Description*

The core of the algorithm is the incremental construction of a classification rule of the type

IF<term1 AND term2 AND …> THEN <class>

by an ant. Each term in a rule is attribute-value pair related by an operator. The relation operator we use in our experiment is = sign. An example term is feature with value (f1 = 2). Here the attribute name is f1 and its value is 2. In this paper all continuous attributes are discretized .

The search space has to be defined first to make the ant to move and find the solution. Here the dataset used for classification forms the search space. The features of the search space are the attributes of the data set. For example, a feature called 'f1' may have four possible values {1, 2, 3, 4}. The task of the ant is to visit a feature and choose one of its possible values to form a term in the antecedent condition of a rule. When a feature has been visited, it cannot be visited again by an ant, because the condition part of the type (f1 = 1 or f1 = 2) is not permitted. The ants may visit the possible features in any order and may not visit some features at all. The search space can be represented by the graph as follows.

The general description of Ant-Miner algorithm given by Parepinelli (2002)[11] is shown in Figure1. Ant-Miner-T is a sequential covering algorithm with genetic optimization of some parameters. It discovers a rule and the training samples correctly covered by this are removed from the training set. The algorithm discovers another rule using the reduced training set and after its discovery the training set is further reduced by removing the training samples covered by the newly discovered rule. This process continues until the training set is almost empty or the training set cannot be reduced further. In the original Ant-Miner algorithm the process is repeated until the training set is almost empty or the user defined threshold for terminating the algorithm.

B. *Rule Construction*

An important part of the algorithm is the step in which an ant add terms in the antecedent part of the rule that is constructing. The rule construction continues until one term from each attribute has been added with provided each term has minimum number of cases.

C. *Pheromone Initialization*

At the beginning of each iteration of the WHILE loop the pheromone values on the edges between all terms are initialized with the same amount of pheromone. The initial pheromone is

$$\tau_{ij}(t=0) = \frac{1}{\sum_{i=1}^{a} b_i} \qquad (1)$$

where a is the total number of attributes (excluding the class attribute) and bi is the number of possible values that can be taken on by an attribute Ai(A represents the attribute set). Since all the pheromone values are same hence the first ant has no historical information to guide its search.

Algorithm: Ant-Miner **Input : TrainingSet = {all training cases}** **Output : Generated Rules**

TrainingSet = {all training cases};
DiscoveredRuleList = [ ]; /* rule list is initialized with an empty list */
WHILE (TrainingSet remains same)
t = 1; /* ant index */
j = 1; /* convergence test index */
Initialize all trails with the same amount of pheromone;
REPEAT
• Ant t starts with an empty rule and incrementally constructs a classification rule Rt by adding one term at a time to the current rule.
• Prune rule Rt;
• Update the pheromone of all trails by increasing pheromone in the trail followed by Ant t (proportional to the quality of Rt) and decreasing pheromone in the other trails (simulating pheromone evaporation);
IF (Rt is equal to Rt – 1) /* update convergence test */
THEN j = j + 1;
ELSE j = 1;
END IF
t = t + 1
UNTIL (i < No_of_ants) OR (j < No_rules_converg)
Choose the best rule Rbest among all rules Rt constructed by all the ants;
Addrule Rbest to DiscoveredRuleList;
TrainingSet = TrainingSet - {set of cases correctly covered by Rbest};
END WHILE

Figure 1: The Ant-Miner Algorithm

A. *Term Selection*

An ant incrementally adds terms in the antecedent part of the rule that it is constructing. The selection of the next term is subject to the condition that the attribute A_i of that term should not be already present in the current partial rule. In other words, once a term (i.e. an attribute-value pair) has been added in the rule then no other term containing that attribute can be considered. The probability of selection of a term for addition in the current partial rule is given by the following equation:

$$P_{ij} = \frac{\eta_{ij}.^{\alpha}\,\tau_{ij}(t)^{\beta}}{\sum_{i=1}^{a} x_i . \sum_{j=1}^{b_i} (\eta_{ij}{}^{\alpha} . \tau_{ij}(t)^{\beta})} \tag{2}$$

where η_{ij} is the value of a problem-dependent heuristic function for term, the higher the value of η_{ij} the more relevant for classification the term is, and so the higher its probability of being chosen. a is the total number of attributes. xi is set to 1 if the attribute Ai was not yet used by the current ant, or to 0 otherwise bi is the number of values in the domain of the ith attribute. In ACO algorithms it is common to use two parameters called alpha (α) and beta (β) to control the relative importance to the pheromone and heuristic values. In this study it is taken as α = 1and β = 1 after trying genetic optimization algorithm to optimize α and β.

B. *Heuristic Function*

The heuristic value of a term gives an indication of its usefulness and thus provides a basis to guide the search. In traditional ACO, a heuristic value is usually used in conjunction with the pheromone value to decide on the transitions to be made. In Ant-Miner, the heuristic value is taken to be an information theoretic measure for the quality of the term to be added to the rule. Entropy is usually used to describe the information contained in the system. Shannon defined the concept of information as

$$\mathbf{H(W \mid A_i = V_{i,j}) = -\sum_{w=1}^{k} (P(w \mid A_i = V_{i,j}) \bullet \log_2 (P(w \mid A_i = V_{i.j})} \tag{3}$$

This type of entropy calculation is used in the original Ant-Miner. After experimenting with Non-Shannon entropy measures, it is found that Tsallis entropy produces significant results. Hence we propose Tsallis entropy based Ant-Miner called Ant-Miner-T as follows:

$$H(W \mid A_i = V_{ij}) = \frac{1 - \sum_{w=1}^{k} p^{\alpha}(W \mid A_i = V_{ij})}{\alpha - 1} \tag{4}$$

where $\alpha > 0$ and $\alpha \neq 1$ is a parameter which is greater than 0. The Tsallis entropy is a generalization of the standard Boltzmann-Gibbs entropy. It was an extension put forward by Constantino Tsallis in 1988. In this case, p denotes the probability distribution of interest, and q is a real parameter. In the limit as $\alpha \rightarrow 1$, the normal Boltzmann-Gibbs entropy is recovered.

D. Rule Quality and Pruning

1) Quality of a Rule

The quality of a rule, denoted by *Q*, is computed by the formula:

$$Q = sensitivity * specificity \tag{5}$$

This can be defined as

$$Q = \frac{TP}{TP + FN} * \frac{TN}{FP + TN} \tag{6}$$

- TP (True Positives) is the number of cases covered by the rule that have the class predicted by the rule.
- FP (False Positives) is the number of cases covered by the rule that have a class different from the class predicted by the rule.
- FN (False Negatives) is the number of cases that are not covered by the rule but that have the class predicted by the rule.
- TN (True Negatives) is the number of cases that are not covered by the rule and that do not have the class predicted by the rule.

Q´s value is within the range $0 \leq Q \leq 1$ and, the larger the value of *Q*, the higher the quality of the rule.

E. Rule Pruning

Rule pruning is a common technique in data mining. The main goal of rule pruning is to remove irrelevant terms that might have been unduly included in the rule. Rule pruning potentially increase the predictive power of the rule, helping to avoid its over fitting to the training data. Another aspect of rule pruning is that it improves the simplicity of the rule, since a shorter rule is usually easier to be understood by the user than a longer one.

F. Pheromone Updating

The pheromone values are updated so that the next ant can make use of this information in its search. The amount of pheromone on each term occurring in the rule is updated according to the equation:

$$\tau_{ij}(t+1) = \tau_{ij}(t) + \tau_{ij}(t).Q \quad \forall i, j \in R \tag{7}$$

where R is the set of terms occurring in the rule constructed by the ant at iteration t. In Ant-Miner, pheromone evaporation is implemented in an indirect way. More precisely, the effect of pheromone evaporation for unused terms is achieved by normalizing the value of each pheromone T_{ij}. This normalization is performed by dividing the value of each T_{ij} by the summation of all T_{ij}.

IV. NON -SHANNON ENTROPY MEASURES

In this section, the proposed ACO algorithm based on Tsallis entropy is detailed for mammogram region classification.

G. Non-Extensiveness in Mammograms

Mammograms are more difficult to interpret when the breast tissues are dense in nature. Breast tissue is composed of non-dense tissue (fat) and dense tissue (glands, ligaments and stromal tissue) and pectoral muscle. Dense breast tissue appears as a solid white area on a mammogram and fat appears as a dark area. Abnormalities in mammogram are also dense tissue and appear as solid white areas. This makes mammograms highly fractal and difficult to analyze. Non-Extensiveness concept enables researchers to find a consistent treatment of dynamics in many Non-Extensive physical systems such as long-range interactions, long-time memories, and multi-fractal structures, which cannot be explained within the Boltzmann Gibbs (BG) statistics (M. Portes de Albuquerque 2004)[19] conducted tests to check how good the Non-Extensive entropic thresholding is for different classes of images, and also to analyze the influence of Tsallis parameter α in the segmentation result. The results show that the Tsallis entropy (TE) is performing well when the system is Non-Extensive and fractal.

H. Shannon vs. Non-Shannon Entropy:

From a conventional point of view, the entropy is a basic thermodynamic concept that is linked with the order of irreversible processes in the universe. Physically it can be associated with the amount of disorder in a physical system. The Shannon Entropy (SE) may be system that obey Bolzman Gibbs statistics are called extensive systems. If we consider that a physical system can be merged into two statistical independent subsystems A and B, the probability of the composite system is P(A + B) = P(A) + P(B). It has been proved that the SE has the extensive property (additivity): S(A + B) = S (A) + S(B). There are certain classes of physical systems like mammograms, which entail long-range interactions, long time memory and fractal-type structures; so definitely a kind of extension appears to become necessary with the existing model. The Non-Extensive entropy is a recent development in statistical mechanics and it is a new formalism in which a real quantity q was introduced as parameter for physical systems that present long range interactions, long time memories and fractal-type structures.

I. Shannon Entropy

The concept of Shannon entropy (C.E. Shannon 1948)[20] is the central role of information theory sometimes referred as measure of uncertainty. The entropy of a random variable is defined in terms of its probability distribution and can be shown to be a good measure of randomness or uncertainty.

$$S = -\sum_{i=1}^{k} H_i \log_2 H_i \tag{8}$$

J. Non -Shannon Entropy Measures

1) Renyi Entropy

In information theory, the Rényi entropy, a generalisation of Shannon entropy, is one of a family of functional for quantifying the diversity, uncertainty or randomness of a system (Rényi A. 1960)[21]. It is named after Alfréd Rényi.

$$R = \frac{1}{1-\alpha}\log_2\left(\sum_{i=1}^{k} H_i^{\alpha}\right) \quad \text{here} \quad \alpha \neq 1, \alpha > 0 \tag{9}$$

K. Havrda and Charvat Entropy

A well-known generalization of the Shannon entropy is the Havrda and Charvat entropy of order α is a strictly concave function of the probability distribution and satisfies the decisivity and maximality properties. (with the exception that its maximal value is ln n only if α = 1)(Havrda J. et al. 1967)[22].

$$HC = \frac{1}{1-\alpha}\left(\sum_{i=1}^{k} H_i^{\alpha} - 1\right) \text{ here} \quad \alpha \neq 1, \alpha > 0 \tag{10}$$

As the Shannon entropy, the Havrda and Charvat entropy of order α is a strictly concave function of the probability distribution and satisfies the decisivity and maximality properties. (with the exception that its maximal value is ln n only if α = 1).

L. Tsallis Entropy

Tsallis proposed to replace the usual Gibbs extensive entropy with his Non-Extensive entropy, and maximize that, subject to some constraints. He got an infinite family of Tsallis Non-Extensive entropies, indexed by α (actually he called it “q”), which quantifies the degree of departure from extensivity. One can get back the Gibbs entropy by making $\alpha \rightarrow 1$. This Non-Extensive entropy is exactly the same as Havrda-Charvat structural α-entropy is hugely neglected by the Non-Extensive mechanics community (Portes de Albuquerque M. et al. 2004)[19].

$$T = \frac{1}{\alpha-1}\left(1 - \sum_{i=1}^{k} H_i^{\alpha}\right) \quad \text{here} \quad \alpha \neq 1, \alpha > 0 \tag{11}$$

Kapur's Entropy (J. N. Kapur 1994)[23]

$$K_{\alpha,\beta} = \frac{1}{\beta - \alpha} \log_2 \frac{\sum_{i=1}^{k} H_i^{\alpha}}{\sum_{i=1}^{k} H_i^{\beta}}$$

here $\alpha \neq 1, \alpha > 0, \beta > 0$ (12)

V. EXPERIMENTAL SETUP

In this section, experimental set up is performed for mammogram region classification as normal, benign and malignant. Ant-Miner is one of the classification algorithm uses the Shannon entropy originated from the information theory. Non-Shannon entropy are applied as general entropy formalism for information theory. For the first time Ant-Miner classification based on Non-Extensive entropy is proposed regarding the presence of non-additive information content in some mammogram classes. In Non-Shannon entropy a new parameter and α are β introduced as real numbers associated with the Non-Extensivity of the system, and it is system dependent.

The following parameters used to achieve the results:

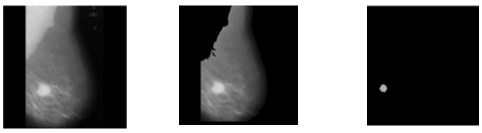
Number of Ants = 100
Minimum cases per rule = 10
Number of Rule Converge = 10

In our experiments, Mini-MIAS database has been used. It is available a http://peipa.essex.ac.uk.

M. Image Segmentation

The preprocessing of mammogram image is essential before detection and segmentation of microcalcification. However, the presence of artifacts and pectoral muscle can disturb the detection of microcalcification and reduce the rate of accuracy in the Computer Aided Diagnosis (CAD). Its inclusion can affect the results of intensity-based image processing methods and needs to be identified and removed before further analysis. These processes are performed in the preprocessing stage (Velayutham C. et al. 2004)[24]

Image segmentation is one of the most critical tasks in automatic image analysis. Segmentation consists of subdividing an image into its constituent part and extracting those of interest. Many techniques for global thresholding have been developed over the years to segment images and recognize patterns but the error on the segmentation leads to misclassification. In this study mammogram is segmented using rough set theory (Thangavel K. et al.)[25]. The original and the segmented images are shown in Figure 2.



a. Origional b. After Pre-Processing c. After Segmentation

Figure 2: Mammogram Segmentation

N. Feature Extraction

Since the classification algorithm requires the classified data to be composed of feature vectors, data mining cannot be directly performed on the original image. The Gray Level Co-occurrence Matrix (GLCM) is a well-established robust statistical tool for extracting second order texture information from images (Dougherty J. et al. 1995)[26]. The GLCM characterizes the spatial distribution of grey levels in an image. Specifically, an element in the GLCM, Pd,θ (i , j), represents the probability of occurrence of the pair of grey levels (i , j) separated by a distance d at direction θ. In this paper, four GLCMs are computed, corresponding to four different directions (θ = 0°, 45°, 90°, 135°) with one distance (d = 1 pixel). The 14 Haralick features are derived from each GLCM: Angular second moment(f1), Contrast(f2), Correlation(f3), Variance(f4), Inverse second different moment(f5), Sum Average(f6), Sum Variance(f7), Sum Entropy(f8), Entropy(f9), Difference Variance(f10), Difference Entropy(f11), Measure of Correlation 1(f12), Measure of Correlation 2 (f13), and Local Mean(f14).

O. GLCM Construction

GLCM is a matrix S that contains the relative frequencies with which two pixels one with gray level value i and the other with gray level j - separated by distance d at a certain angle θ occur in the image. Given an image window W(x, y, c), for each discrete values of d and θ the GLCM matrix S(i, j, d, θ) is defined as follows:

An entry in the matrix S gives the number of times gray level i is oriented with respect to gray level j such that where W(x1, y1) = i and W(x2, y2) = j then

$$(x2, y2) = (x1, y1) + (d * \cos(\theta), d * \sin(\theta))$$

We use distance d = 1 for four different angles θ = {0^o, 45^o, 90^o, 135^o}. Here, angle representation is taken in clock wise direction. For instance consider the following intensity matrix:

$$\begin{pmatrix} 1 & 3 & 1 & 1 & 1 \\ 1 & 4 & 1 & 4 & 1 \\ 2 & 2 & 2 & 1 & 1 \\ 1 & 1 & 2 & 2 & 1 \end{pmatrix}$$

Different intensity values are 1, 2, 3 and 4.

GLCM	
Degree	***Degree = 45o*** ***Distance = 1*** ***1 2 3 4***
1	3 3 0 0
2	3 2 0 1
3	0 0 0 1
4	1 1 0 1

P. Discretization

A discretization algorithm is applied in order to handle problems with real-valued attributes with classification. The term “cut-point” refers to a real value within the range of continuous values that divides the range into two intervals, one interval is less than or equal to the cutpoint and the other interval is greater than the cut-point. For example, a continuous interval [a, b] is partitioned into [a, c] and (c, b], where c is a cut-point. Cut-point is also known as split-point. For some attributes, if doctors have had existing dividing points, one can adopt it directly. For example, patients’ weight can be divided to thin, common and heavy; their age can be divided into under age, youth and the elderly; Medical test results can be said to be normal and abnormal. But for extracted image feature value, there is no existing threshold. In (Chan et al. 2006)[27] the discretization using the k-means algorithm is presented in detail.

In the implementation, two dimensional arrays are used to represent the attribute and the possible values it takes. Each row corresponding to the feature value and the column corresponding to the each possible value it takes. Hence there are only fourteen rows corresponding to each feature. The column corresponding to the maximum possible value an attribute can take.

VI. EXPERIMENTAL RESULTS

All the non-Shannon Entropy measures described in section 4 are used in the classification task as the heuristic functions and the results are reported. These entropies use the parameters α in all the non Shannon entropy measure discussed above and β in the Kapur’s Entropy. It is stated that α > 0, α ≠ 1 and β > 0. It is assumed that β takes the value 1.5 and it is experimented for different values of α such as 0.5 , 2, 3, 4, 5 and 6. The results of classification accuracy for α = 6 and β = 1.5 is reported in Tables 1 and the graphical representation is shown in Figure 3.

TABLE I : CLASSIFICATION ACCURACY OF NON-SHANNON ENTROPY MEASURES AT A = 6, B = 1.5

Data Set (Angle)	Renyi Entropy (%)	Havrda and Charavat Entropy (%)	Tsallis Entropy (%)	Kapur's Entropy (%)
0^o	75.60	86.30	87.70	74.60
45^o	73.30	84.60	89.10	77.50
90^o	78.30	91.60	94.40	80.70
135^o	76.80	90.00	90.30	78.70

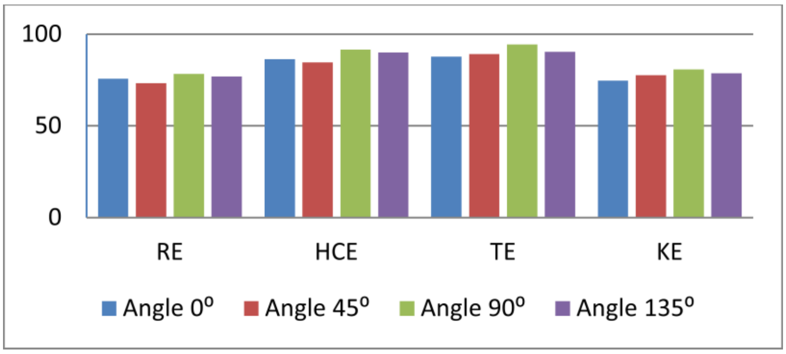


Figure 3: Classification Accuracy of Non-Shannon Entropy Measures at $\alpha = 6$, $\beta = 1.5$

In the above figure RE represents Renyi Entropy, HCE represents Havrda and Charavat Entropy, TE represents Tsallis Entropy and KE represents Kapur's Entropy.

It is observed from experimental resultsthat Tsallis Entropy and Havrda and Charava Entropy produce better result in all the cases comparing to other non-shannon entropy measures.

The classification accuracy of Havrda and Charavat Entropy and the Tsallis Entropy is presented in Table 2 and is plotted in Figure 5. It is concluded from the Table 2 and Figure 4 that Tsallis Entropy produces better results comparing to Havrda and Charavat Entropy in all Angles.

TABLE II: CLASSIFICATION ACCURACY OF HAVRDA ANDCHARAVAT ENTROPY VS. TSALLIS ENTROPY

Data Set (Angle)	**Havrda and Charavat Entropy**	**Tsallis Entropy**
0°	86.40	89.20
45°	88.30	90.90
90°	95.80	96.10
135°	90.00	90.80

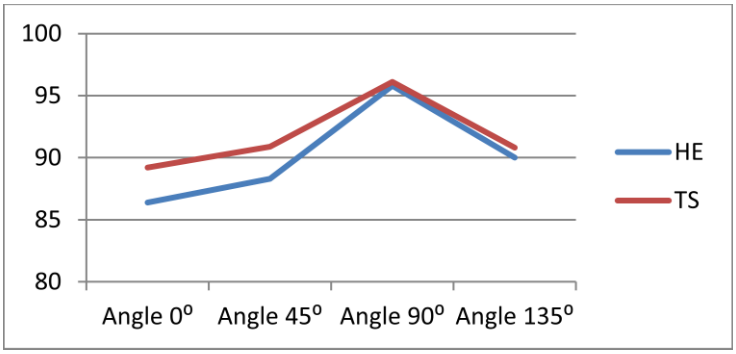


Figure 4: Classification accuracy of Havrda and Charavat Entropy vs. Tsallis Entropy

VII. COMPARATIVE ANALYSIS

The predictive accuracies of mammogram features at angles 0^o, 45^o, 90^o and 135^o using C4.5 and Ant-Miner and Tsallis Ant-Miner are listed in Table 3 and are plotted in Figure 4.

TABLE III: REPORT OF CLASSIFICATION ACCURACY USING C4.5, ANT-MINER AND ANT-MINER-T

Data Set (Angle)	**C4.5 (%)**	**Ant-Miner (%)**	**Ant-Miner-T (%)**
0°	90.80	88.60	89.20
45°	90.30	90.40	90.90
90°	95.20	94.80	96.10
135°	88.80	90.10	90.80

The results indicate that the Ant-Miner using Tsallis entropy achieves little higher accuracy rate in mammogram feature set at Angle 0^o, Angle 45^o and Angle 90^o and Angle 135^o.

TABLE IV: NUMBER OF RULES, TPR AND FPR (ANT-MINER VS. ANT-MINER-T)

Data Set	**No. of Rules**			**TPR**			**FPR**		
	C4.5	***AM***	***AM-T***	***C4.5***	***AM***	***AM-T***	***C4.5***	***AM***	***AM-T***
Angle 0°	19.60	12.20	11.00	00.95	00.83	00.81	00.25	00.09	00.09
Angle 45°	21.00	10.00	09.90	00.94	00.94	00.95	00.03	00.03	00.01
Angle 90°	19.30	12.60	12.10	00.89	00.92	00.94	00.06	00.04	00.03
Angle 135°	10.30	12.50	11.30	00.83	00.85	00.86	00.08	00.07	00.06

The number of rules generated, TPR and FPR for mammogram features at angles 0^o, 45^o, 90^o and 135^o using C4.5 and Ant-Miner and Tsallis Ant-Miner are listed in Table 4.

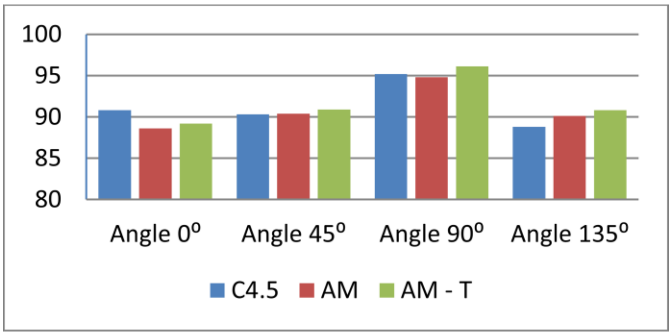


Figure 5: Classification accuracy using C4.5, Ant-Miner and Ant-Miner-T

The number of rules generated by Ant-Miner-T is less compared to C4.5, Ant-Miner. Significant improvement is shown in the TPR for the features extracted at Angle 45^o, Angle 90^o and Angle 135^o. In all angles the false positive rate becomes better when compared to other classifiers.

As quoted by (Rathi et al. 2008) the Tsallis nonextensive entropy of the statistical physics literature exactly matches the previously defined Havrda-Charvat structural α-entropy of information theory, Tsallis entropy based Ant-Miner reports the approximately the same accuracy as reported by Kavrda and Charavat entropy. Since in the literature there is evidence for performance improvement using Tsallis entropy this paper proposes Ant-Miner-T which uses Tsallis entropy as its heuristic function.

VIII. CONCLUSION

This paper analysed the performance of different non-shannon entropy measures as the heuristic function of Ant-Miner. Different values of the parameter α are taken into account. It is concluded that Tsallis Entropy based Ant-Miner (Ant-Miner-T) produces better results in the features extracted in all angles viz. Angle 0^o, Angle 45^o, Angle 90^o, and Angle 135^o when compared with Ant-Miner. A comparative analysis is also performed with C4.5, Ant-Miner and Ant-Miner-T. Comparative study reveals that Ant-Miner-T outperforms Ant-Miner in all Angles but it is unable to produce better accuracy at Angle 0^o when compared with C4.5. The number of rules generated is less and the TPR and FPR are better when compared with other classifiers. The heuristic function which uses Tsallis entropy does not require logarithmetic calculation as it is needed in Shannon Entropy.

REFERENCES

[1] C. Varela, P. G. Tahoces, A.J. Méndez, M. Souto, & J.J. Vidal, "Computerized Detection of Breast Masses in Digitized Mammograms. Computers in Biology and Medicine," 37, pp. 214–226, 2007.

[2] F. Fauci, S. Bagnasco, R. Bellotti, D. Cascio, S.C. Cheran, F. De Carlo, G. De Nunzio, M.E. Fantacci, G. Forni, Lauria, A., Torres, E.L., Magro, R., Masala, G.L., Oliva, P., Quarta, M., Raso, G., Retico, A. & Tangaro, S.(2004). "Mammogram Segmentation by Contour Searching and Massive Lesion Classification with Neural Network," Proc. IEEE Nuclear Science Symposium Conference Record, Rome, Italy, IEEE Press, 5, pp. 2695–2699, DOI: 10.1109/NSSMIC.2004.1462823, 2004.

[3] Jiawei Han & Micheline Kamber, "Data Mining Concepts and Techniques," Morgan Kaufmann, San Francisco, USA, 2006.

[4] E. Bonabeau, M. Dorigo, & G. Theraulaz, "Swarm Intelligence: From Natural to Artificial Systems," New York, Oxford University Press, 1999.

[5] M. Dorigo, A. Colorni, & V. Maniezzo, "The Ant System: optimization by a colony of cooperating agents," IEEE Transactions on Systems, Man, and Cybernetics-Part B, 26, pp. 29-41, 1996.

[6] H. Li, M.L. Giger, Y. Yuan, L. Lan, & C.A. Sennett, "Performance of CADx on a Large Clinical Database of FFDM Images," Proc. IWDM, LNCS, pp. 510–514, DOI : 10.1007/978-3-540-70538-3-71, 2008.

[7] J. Jiang, B.Yao, & A.M. Wason, "A genetic algorithm design for microcalcification detection and classification in digital mammograms," Computerized Medical Imaging and Graphics, 31, pp.49–61, 2007.

[8] D. Stylianos, E. Tzikopoulos, Michael, V. Mavroforakis, Harris, Georgiou, Nikos Dimitropoulos & Sergios Theodoridis , "A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry," Computer Methods and Programs in Biomedicine, 102 , pp. 47–63, 2011.

[9] Jinchang Ren., "ANN vs. SVM: Which one performs better in classification of MCCs in mammogram imaging," Knowledge-Based Systems, 26, pp. 144–153, 2012.

[10] F. Juan, Ramirez, Villegas & F. David, Ramirez, Moreno, "Wavelet packet energy, Tsallis entropy and statistical parameterization for support vector-based and neural-based classification of mammographic regions," Neurocomputing, 77, pp. 82-100, 2012.

[11] R.S. Parepinelli, H.S. Lopes, & A. Freitas, "An Ant Colony Algorithm for Classification Rule Discovery". Proc. H. A. a. R. S. a. C. Newton, Data Mining: Heuristic Approach, Idea Group Publishing, pp. 31- 35, 2002.

[12] Bo, Liu, A. Hussien, Abbass & Bob Mckay, "Density-Based Heuristic for Rule Discovery with Ant-Miner, In the joint Workshop of Australia-Japan on intelligent and evolutionary system," pp. 180 – 184, 2002.

[13] Bo, Liu, A. Hussein, Abbass & Bob McKay, "Classification Rule Discovery with Ant Colony Optimization," IEEE Computational Intelligence Bulletin, 3, pp. 31-35, 2004.

[14] David, Martens, Manu, De, Backer, Raf, Haesen, Jan, Vanthienen, Monique, Snoeck & Bart, Baesens, "Classification with Ant Colony Optimization. IEEE Transactions on Evolutionary Computation," 11, pp 651-665, 2007.

[15] J. Smaldon, & A.A. Freitas, "A New Version of the Ant-Miner Algorithm Discovering Unordered Rule Sets," Proc. Genetic and Evolutionary Computation Conference, Seattle, Washington, pp. 43 -50, DOI. 10.1145/1143997.1144004, 2006.

[16] N. Holden, and. A.A. Freitas, "A hybrid PSO/ACO algorithm for classification," Proc. GECCO Workshop on Particle Swarms: The Second Decade, ACM Press, New York ,pp. 1- 11, 2007.

[17] Swaminathan, "Rule Induction using Ant colony Optimization for Mixed Variable Attributes," MSc Thesis, Texas Tech. Univ, 2006.

[18] E. Fernando, B. Otero, A. Alex, Freitas & G. Colin, Johnson, "cAnt-Miner: An Ant Colony Classification Algorithm to Cope with Continuous Attributes, Ant Colony Optimization and Swarm Intelligence," Lecture Notes in Computer Science , 5217, pp. 48-59, 2008.

[19] M. Portes de Albuquerque, I.A. Esque, & A.R. Gesualdi Mello, "Image thresholding using Tsallis entropy. Pattern Recognition Letters," 25, pp. 1059-1065, 2004.

[20] C.E. Shannon, (1948). "A Mathematical Theory of Communication," Bell System Technical Journal, 27, pp. 379-423, 1948.

[21] A. Rényi "On measures of information and entropy," Proc. 4th Berkeley Symposium on Mathematics, Statistics and Probability, pp. 547–561, 1960.

[22] J. Havrda and F. Charvat "Quantification method in classification processes: concept of structural α entropy," Kybernetika, 3, pp 30-35, 1967.

[23] J.N. Kapur, "Measure of information and their applications," 1st edition Wiley Eastern Limited, New Delhi, 1994.

[24] C. Velayutham, & K. Thangavel, "Detection and Elimination of Pectoral Muscle in Mammogram Images using Rough Set Theory," Proc. International Conference on advances in Engineering, Science and Management, Nagapattinam, pp. 48-54, 2012.

[25] K. Thangavel, & C. Velayutham, "Segmentation Using Rough Set Theory," International Journal Computational Intelligence and Informatics, pp. 255-261, 2012.

[26] Chan & A. Freitas, "A new ant colony algorithm for multi-label classification with applications in bioinformatics,". Proc. Genetic and Evolutionary Computation Conference, USA, pp. 27-34, 2006.

[27] J. Dougherty, R. Kohavi, & M. Sahami, "Supervised and unsupervised discretization of continuous features," Proc. 12th International Conference on Machine Learning, Morgan Kaufmann San Francisco, pp. 194-202, 1995.