CHAPTER - 5

CLASSIFICATION OF MAMMOGRAM: TSALLIS ENTROPY BASED ANT-MINER

5.1 INTRODUCTION

In the original Ant-Miner, the heuristic function which is used by the ants for a selection of a path uses Shannon Entropy (SE) measure. In this chapter, a novel idea of using non-Shannon Entropy measure is discussed and finally it is reported that Tsallis entropy based Ant-Miner gives better result comparing to other non-Shannon Entropy measures. Moreover Tsallis entropy does not require the computations of logarithm as in the case of Ant-Miner.

The rest of the chapter is organized as follows: Section 5.2 explains Entropy as an information measure; Section 5.3 deals with Shannon entropy measure; Section 5.4 describes non-Shannon entropy measures; Section 5.5 states the heuristic functions in Artificial Intelligence (AI); Section 5.6 elucidates heuristic function in Ant-Miner; Section 5.7 gives out the experimental results, and in Section 5.8 a comparative analysis is performed and this Chapter is concluded in Section 5.9.

5.2 ENTROPY AS AN INFORMATION MEASURE

In information theory, Entropy is a measure of the uncertainty associated with a random variable. It is also a measure of disorder, or more precisely unpredictability. For example, a series of coin tosses with a fair coin has maximum entropy, since there is no way to
predict what will come next. A string of coin tosses with a coin with two heads and no tails has zero entropy, since the coin will always come up heads.

5.3 SHANNON ENTROPY MEASURE

This concept was introduced by Claude E. Shannon in 1948 in a paper "A Mathematical Theory of Communication" [130]. The SE may be described as:

\[ - \sum_{i=1}^{k} p_i \log_2 p_i \]

Systems that obey Boltzmann Gibbs (BG) statistics are called extensive systems. If it is considered that a physical system can be merged into two statistical independent subsystems A and B, the probability of the composite system is \( p^{A+B} = p^A \cdot p^B \) and it has been proved that the SE has the extensive property (additivity):

\[ S(A+B) = S(A) + S(B) \]

5.4 NON-SHANNON ENTROPY MEASURES

In most feature descriptors, Shannon’s measure is used to measure entropy. In this analysis, non-Shannon measures are used to measure entropy which is used as a heuristic function in Ant-Miner. Non-Shannon entropies have a higher dynamic range than Shannon entropy over a range of scattering conditions, and are therefore useful in estimating scatter density and regularity [132, 133].

5.4.1 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 mammogram highly fractal and difficult to analyze. Non-extensiveness concept enables researchers to find a consistent treatment of dynamics in many nonextensive 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 [20] conducted tests to check how good the nonextensive entropic thresholding is for different classes of images, and also to analyze the influence of Tsallis parameter $\alpha$ in the segmentation result. The results show that the Tsallis entropy (TE) is performing well when the system is non-extensive and fractal.

5.4.2 Shannon vs. Non-Shannon Entropy

The Shannon Entropy (SE) may be systems that obey Boltzman Gibbs statistics which are called extensive systems. 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[134].
Rényi 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 [135]. It is named after Alfréd Rényi.

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

Havrda and Charvat Entropy

A well-known generalization of the Shannon Entropy is the Havrda and Charvat Entropy of order $\alpha$ is a strictly concave function of the probability distribution and satisfies the decisivity and maximality properties. (with the exception that its maximal value is $\log_2 n$ only if $\alpha = 1$). [136]

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

As the Shannon Entropy, the Havrda and Charvat Entropy of order $\alpha$ is a strictly concave function of the probability distribution and satisfies the decisivity and maximality properties. (with the exception that its maximal value is $\log_2(n)$ only if $\alpha = 1$).

Tsallis Entropy

Tsallis is proposed to replace the usual Gibbs extensive entropy with his nonextensive entropy, and maximize that, subject to some constraints. He got an infinite family of Tsallis nonextensive entropies, indexed by $\alpha$ (actually he called it “$q$”), which quantifies the degree of departure from extensivity. One can get back the Gibbs
entropy by making $\alpha \to 1$. This non-extensive entropy is exactly the same as Havrda-Charvát structural $\alpha$-entropy is hugely neglected by the non-extensive mechanics community [20]

$$T = \frac{1}{\alpha} \left( -\sum_{i=1}^{k} p_i^\alpha \right) \quad \text{here} \quad \alpha = 1$$

**Kapur’s Entropy** [137]

$$K_{\alpha, \gamma} = \frac{1}{\beta - \alpha} \log_2 \left( \sum_{i=1}^{k} \frac{p_i^{\alpha}}{\sum_{i=1}^{k} p_i^{\beta}} \right) \quad \text{here} \quad \alpha = 0, \beta > 1$$

### 5.5 HEURISTIC FUNCTIONS IN ARTIFICIAL INTELLIGENCE

A Heuristic technique helps in solving problems. There are heuristics of every general applicability as well as domain specific. The strategies are general purpose heuristics. In order to use them in a specific domain they are coupled with some domain specific heuristics. There are two major ways in which domain-specific, heuristic information can be incorporated into rule-based search procedure.

- In the rules themselves
- As a heuristic function that evaluates individual problem states and determines how desired they are.

A heuristic function is a function that maps from problem state description to measures desirability, usually represented as number weights. The value of a heuristic function at a given node in the search process gives a good estimate of that node being on the desired path to solution. Well designed heuristic functions can provide a
fairly good estimate of whether a path is good or not. The purpose of a heuristic function is to guide the search process in the most profitable directions, by suggesting which path to follow first when more than one path is available. However in many problems, the cost of computing the value of a heuristic function would be more than the effort saved in the search process. Hence, generally there is a trade-off between the cost of evaluating a heuristic function and the savings in search that the function provides.

A heuristic function ranks alternatives in various search algorithms at each branching step based on the available information (heuristically) in order to make a decision about which branch to follow during a search.

5.6 HEURISTIC FUNCTION IN ANT-MINER

As discussed in the previous chapter, Ant-Miner searches the problem space for adding a term into the rule using the function given in Chapter 4, Equation (probability calculation for selecting a term). This equation uses the pheromone value and the problem dependent heuristic function. The heuristic value calculation is given by the formula:

\[ H(W | F_i = \zeta_{i,j}) = \sum_{w} (W | F_i = \zeta_{i,j}) \cdot \log_2(P(W | F_i = \zeta_{i,j})) \]

where \( W \) is the class attribute (i.e., the attribute whose domain consists of the classes to be predicted) and \( k \) is the number of classes. \( P(w | F_i = V_{ij}) \) is the empirical probability of observing class \( w \) conditional on having observed \( F_i = V_{ij} \). The higher the value of \( H(W | F_i = V_{ij}) \), the more uniformly distributed the classes are and so,
the smaller the probability that the current ant chooses to add term\(_{ij}\) to its partial rule. The information-theoretic heuristic function is given in Chapter 4.

### 5.6.1 Representation of Problem Space

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. Initial Pheromone value for Angle 90\(^0\), Fold 1 training set is given by the below matrix. Here the zero value corresponding to null (i.e.) feature 2 (\(f_2\)) has only seven possible values, feature 3 (\(f_3\)) has only six possible values and so on. Maximum feature value an attribute may take is 9. For example feature 8 (\(f_8\)), feature 9(\(f_9\)) and feature 13(\(f_{13}\)) take nine possible values.

\[
\begin{bmatrix}
0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0000 \\
0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0000 & 0.0000 \\
0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0000 & 0.0000 \\
0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0000 & 0.0000 \\
0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0000 & 0.0000 \\
0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0000 & 0.0000 \\
0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0000 & 0.0000 \\
0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0000 & 0.0000 \\
0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0000 & 0.0000 \\
0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0000 & 0.0000 \\
0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0000 & 0.0000 \\
0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0000 & 0.0000 \\
0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0000 & 0.0000 \\
0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0000 & 0.0000 \\
0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0000 & 0.0000 \\
0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0000 & 0.0000 \\
0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0000 & 0.0000 \\
0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0000 & 0.0000 \\
0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0000 & 0.0000 \\
0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0000 & 0.0000 \\
0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0000 & 0.0000 \\
0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0101 & 0.0000 & 0.0000
\end{bmatrix}
\]
The entropy value is given by the following matrix

\[
\begin{bmatrix}
0.0000 & 0.0000 & 1.2216 & 0.6801 & 1.0000 & 1.1246 & 0.9496 & 0.7219 & 0.0000 \\
0.0000 & 0.3451 & 0.0000 & 0.4855 & 1.5000 & 1.1779 & 0.0000 & 0.0000 & 0.0000 \\
1.5751 & 0.6784 & 0.5548 & 1.3431 & 0.9893 & 1.1683 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 1.1635 & 0.7496 & 1.2478 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 1.2216 & 0.8905 & 0.8791 & 1.3269 & 1.3922 & 0.3451 & 0.0000 & 0.0000 \\
0.0000 & 0.5033 & 1.3522 & 1.0736 & 0.9341 & 1.1683 & 0.0000 & 0.0000 & 0.0000 \\
0.7219 & 1.2076 & 1.0877 & 1.1707 & 0.0000 & 0.0000 & 1.2216 & 0.0000 & 0.0000 \\
0.7219 & 1.2076 & 1.0877 & 1.1707 & 0.0000 & 0.0000 & 1.2216 & 0.0000 & 0.0000 \\
1.2846 & 1.2260 & 0.8905 & 0.9968 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
1.2375 & 0.9740 & 1.2260 & 0.0000 & 0.9457 & 0.7219 & 0.0000 & 0.0000 & 0.0000 \\
0.8813 & 0.8213 & 0.8204 & 0.9518 & 1.0724 & 0.8315 & 0.0000 & 0.0000 & 0.0000 \\
0.7219 & 0.5436 & 1.1463 & 0.9877 & 0.8728 & 0.0000 & 0.6421 & 0.6790 & 0.0000 \\
0.9044 & 1.1712 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
\end{bmatrix}
\]

The heuristic value is given by the following matrix

\[
\begin{bmatrix}
0.0173 & 0.0173 & 0.0040 & 0.0099 & 0.0064 & 0.0050 & 0.0070 & 0.0094 & 0.0000 \\
0.0173 & 0.0136 & 0.0173 & 0.0120 & 0.0009 & 0.0045 & 0.0173 & 0.0000 & 0.0000 \\
0.0001 & 0.0099 & 0.0113 & 0.0026 & 0.0065 & 0.0046 & 0.0000 & 0.0000 & 0.0000 \\
0.0173 & 0.0173 & 0.0173 & 0.0046 & 0.0091 & 0.0037 & 0.0173 & 0.0000 & 0.0000 \\
0.0173 & 0.0040 & 0.0076 & 0.0077 & 0.0028 & 0.0021 & 0.0136 & 0.0000 & 0.0000 \\
0.0173 & 0.0118 & 0.0025 & 0.0056 & 0.0071 & 0.0173 & 0.0000 & 0.0000 & 0.0000 \\
0.0173 & 0.0173 & 0.0173 & 0.0046 & 0.0042 & 0.0040 & 0.0173 & 0.0000 & 0.0000 \\
0.0094 & 0.0041 & 0.0054 & 0.0045 & 0.0173 & 0.0173 & 0.0040 & 0.0173 & 0.0173 \\
0.0094 & 0.0041 & 0.0054 & 0.0045 & 0.0173 & 0.0173 & 0.0040 & 0.0173 & 0.0173 \\
0.0033 & 0.0039 & 0.0076 & 0.0064 & 0.0173 & 0.0173 & 0.0000 & 0.0000 & 0.0000 \\
0.0038 & 0.0067 & 0.0039 & 0.0173 & 0.0070 & 0.0094 & 0.0173 & 0.0000 & 0.0000 \\
0.0077 & 0.0084 & 0.0084 & 0.0069 & 0.0056 & 0.0082 & 0.0173 & 0.0000 & 0.0000 \\
0.0094 & 0.0114 & 0.0048 & 0.0065 & 0.0078 & 0.0173 & 0.0103 & 0.0099 & 0.0173 \\
0.0074 & 0.0045 & 0.0173 & 0.0173 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
\end{bmatrix}
\]
5.6.2 Rule Construction

The portion of the Ant-Miner algorithm proposed by Parpinelli et al. [100] where the heuristic function applied is elaborated in Figure 5.1.

\[
\text{Rule} = [] \\
\text{WHILE (usable terms exists) and (all attributes have not been used)} \\
\quad \text{Calculate the heuristic Value,} \\
\quad \text{Generate selection probabilities for each usable term utilizing a} \\
\quad \text{heuristic value and pheromone levels.} \\
\quad \text{Use Roulette wheel selection to select a term T to be added to} \\
\quad \text{the rule.} \\
\quad \text{Add T to Rule.} \\
\text{END WHILE} \\
\text{return Rule}
\]

Figure 5.1: Rule Construction Portion of Ant-Miner

5.7 EXPERIMENTAL RESULTS

All the non-Shannon Entropy measures described in section 5.4 are used in the classification task as the heuristic functions and the results are reported. These entropies use the parameters $\alpha$ in all the non Shannon entropy measure discussed above and $\beta$ in the Kapur’s Entropy. It is stated that $\alpha > 0$, $\alpha \neq 1$ and $\beta > 0$. It is assumed that $\beta$ takes the value 1.5 and it is experimented for different values of $\alpha$
such as 0.5, 1, 2, 3, 4, 5 and 6. The results of classification accuracy are reported in Tables 5.1, 5.2, 5.3, 5.4, 5.5 and 5.6.

**Table 5.1: Classification Accuracy of Non-Shannon Entropy Measures**

at $\alpha = 0.5$, $\beta = 1.5$

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Renyi Entropy (%)</th>
<th>Havrda and Charavat Entropy (%)</th>
<th>Tsallis Entropy (%)</th>
<th>Kapur’s Entropy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angle 0°</td>
<td>74.80</td>
<td>84.00</td>
<td>86.80</td>
<td>77.00</td>
</tr>
<tr>
<td>Angle 45°</td>
<td>77.30</td>
<td>86.70</td>
<td>88.00</td>
<td>75.50</td>
</tr>
<tr>
<td>Angle 90°</td>
<td>79.90</td>
<td>91.70</td>
<td>94.00</td>
<td>80.90</td>
</tr>
<tr>
<td>Angle 135°</td>
<td>77.50</td>
<td>89.40</td>
<td>90.50</td>
<td>78.20</td>
</tr>
</tbody>
</table>

**Figure 5.2: Classification Accuracy of Non-Shannon Entropy Measures**

Measures at $\alpha = 0.5$, $\beta = 1.5$
Table 5.2 Classification Accuracy of Non-Shannon Entropy Measures at $\alpha = 2$, $\beta = 1.5$

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Renyi Entropy (%)</th>
<th>Havrda and Charavat Entropy (%)</th>
<th>Tsallis Entropy (%)</th>
<th>Kapur's Entropy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angle 0°</td>
<td>77.70</td>
<td>86.40</td>
<td>89.20</td>
<td>75.90</td>
</tr>
<tr>
<td>Angle 45°</td>
<td>75.50</td>
<td>88.30</td>
<td>88.70</td>
<td>74.10</td>
</tr>
<tr>
<td>Angle 90°</td>
<td>79.40</td>
<td>95.80</td>
<td>96.00</td>
<td>81.20</td>
</tr>
<tr>
<td>Angle 135°</td>
<td>79.10</td>
<td>89.90</td>
<td>90.40</td>
<td>78.30</td>
</tr>
</tbody>
</table>

Figure 5.3: Classification Accuracy of Non-Shannon Entropy Measures at $\alpha = 2$, $\beta = 1.5$
Table 5.3: Classification Accuracy of Non-Shannon Entropy Measures at $\alpha = 3$, $\beta = 1.5$

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Renyi Entropy (%)</th>
<th>Havrda and Charavat Entropy (%)</th>
<th>Tsallis Entropy (%)</th>
<th>Kapur's Entropy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angle 0°</td>
<td>75.50</td>
<td>83.20</td>
<td>84.80</td>
<td>77.40</td>
</tr>
<tr>
<td>Angle 45°</td>
<td>71.90</td>
<td>86.80</td>
<td>88.10</td>
<td>75.50</td>
</tr>
<tr>
<td>Angle 90°</td>
<td>80.40</td>
<td>93.70</td>
<td>93.70</td>
<td>80.40</td>
</tr>
<tr>
<td>Angle 135°</td>
<td>77.00</td>
<td>88.20</td>
<td>89.90</td>
<td>77.90</td>
</tr>
</tbody>
</table>

Figure 5.4: Classification Accuracy of Non-Shannon Entropy Measures at $\alpha = 3$, $\beta = 1.5$
Table 5.4: Classification Accuracy of Non-Shannon Entropy Measures

at $\alpha = 4, \beta = 1.5$

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Renyi Entropy (%)</th>
<th>Havrda and Charavat Entropy (%)</th>
<th>Tsallis Entropy (%)</th>
<th>Kapur's Entropy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angle 0°</td>
<td>75.30</td>
<td>82.60</td>
<td>87.40</td>
<td>76.10</td>
</tr>
<tr>
<td>Angle 45°</td>
<td>76.90</td>
<td>86.30</td>
<td>90.90</td>
<td>78.60</td>
</tr>
<tr>
<td>Angle 90°</td>
<td>80.50</td>
<td>94.60</td>
<td>96.10</td>
<td>79.40</td>
</tr>
<tr>
<td>Angle 135°</td>
<td>76.00</td>
<td>89.20</td>
<td>90.80</td>
<td>78.30</td>
</tr>
</tbody>
</table>

Figure 5.5: Classification Accuracy of Non-Shannon Entropy Measures at $\alpha = 4, \beta = 1.5$
Table 5.5: Classification Accuracy of Non-Shannon Entropy Measures at $\alpha = 5$, $\beta = 1.5$

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Renyi Entropy (%)</th>
<th>Havrda and Charavat Entropy (%)</th>
<th>Tsallis Entropy (%)</th>
<th>Kapur's Entropy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angle 0°</td>
<td>76.80</td>
<td>86.20</td>
<td>87.10</td>
<td>78.20</td>
</tr>
<tr>
<td>Angle 45°</td>
<td>76.70</td>
<td>89.40</td>
<td>89.70</td>
<td>73.60</td>
</tr>
<tr>
<td>Angle 90°</td>
<td>80.30</td>
<td>93.90</td>
<td>94.10</td>
<td>79.40</td>
</tr>
<tr>
<td>Angle 135°</td>
<td>77.30</td>
<td>88.40</td>
<td>89.80</td>
<td>78.70</td>
</tr>
</tbody>
</table>

Figure 5.6: Classification Accuracy of Non-Shannon Entropy Measures at $\alpha = 5$, $\beta = 1.5$
Table 5.6: Classification Accuracy of Non-Shannon Entropy Measures

at $\alpha = 6$, $\beta = 1.5$

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Renyi Entropy (%)</th>
<th>Havrda and Charavat Entropy (%)</th>
<th>Tsallis Entropy (%)</th>
<th>Kapur's Entropy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angle 0°</td>
<td>75.60</td>
<td>86.30</td>
<td>87.70</td>
<td>74.60</td>
</tr>
<tr>
<td>Angle 45°</td>
<td>73.30</td>
<td>84.60</td>
<td>89.10</td>
<td>77.50</td>
</tr>
<tr>
<td>Angle 90°</td>
<td>78.30</td>
<td>91.60</td>
<td>94.40</td>
<td>80.70</td>
</tr>
<tr>
<td>Angle 135°</td>
<td>76.80</td>
<td>90.00</td>
<td>90.30</td>
<td>78.70</td>
</tr>
</tbody>
</table>

Figure 5.7: Classification Accuracy of Non-Shannon Entropy Measures at $\alpha = 6$, $\beta = 1.5$
5.8 COMPARATIVE ANALYSIS

It is observed from Figures 5.2 to 5.7 that Tsallis Entropy and Havrda and Charava Entropy produce better result in all the cases. The Table 5.7 illustrates the classification accuracy of Havrda and Charavat Entropy at six different values of $\alpha$ (0.5, 1, 2, 3, 4, 5, 6) and these are presented as a graph in Figure 5.8.

Table 5.7: Classification accuracy of Havrda and Charavat Entropy at different values of $\alpha$

<table>
<thead>
<tr>
<th>Data Set</th>
<th>$\alpha = 0.5$</th>
<th>$\alpha = 2$</th>
<th>$\alpha = 3$</th>
<th>$\alpha = 4$</th>
<th>$\alpha = 5$</th>
<th>$\alpha = 6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angle 0°</td>
<td>84.00</td>
<td>83.20</td>
<td>86.40</td>
<td>82.60</td>
<td>86.20</td>
<td>86.30</td>
</tr>
<tr>
<td>Angle 45°</td>
<td>86.70</td>
<td>86.80</td>
<td>88.30</td>
<td>86.60</td>
<td>89.40</td>
<td>84.60</td>
</tr>
<tr>
<td>Angle 90°</td>
<td>91.70</td>
<td>93.70</td>
<td>95.80</td>
<td>94.60</td>
<td>93.90</td>
<td>91.60</td>
</tr>
<tr>
<td>Angle 135°</td>
<td>89.40</td>
<td>88.20</td>
<td>89.90</td>
<td>89.20</td>
<td>88.40</td>
<td>90.00</td>
</tr>
</tbody>
</table>

Figure 5.8: Classification accuracy of Havrda and Charavat Entropy at different values of $\alpha$
It is observed from Table 5.7 and Figure 5.8 that when $\alpha = 3$, Havrda and Charava entropy produces better results at Angle $0^\circ$, Angle $45^\circ$, Angle $90^\circ$ and better results for Angle $135^\circ$ is produced at $\alpha = 6$. In Angle $135^\circ$ the classification accuracy at $\alpha = 3$ is 89.90% and when $\alpha = 6$, it is 90%. Hence it could be concluded that the Havrda and Charavat Entropy produces better results at $\alpha = 3$.

The Table 5.8 illustrates the classification accuracy of Tsallis Entropy at six different values of $\alpha$ (0.5, 1, 2, 3, 4, 5, 6) and it is presented as a graph in Figure 5.9.

### Table 5.8: Classification accuracy of Tsallis Entropy at different values of $\alpha$

<table>
<thead>
<tr>
<th>Data Set</th>
<th>$\alpha = 0.5$</th>
<th>$\alpha = 2$</th>
<th>$\alpha = 3$</th>
<th>$\alpha = 4$</th>
<th>$\alpha = 5$</th>
<th>$\alpha = 6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angle $0^\circ$</td>
<td>86.80</td>
<td>89.20</td>
<td>84.80</td>
<td>87.40</td>
<td>87.10</td>
<td>87.70</td>
</tr>
<tr>
<td>Angle $45^\circ$</td>
<td>88.00</td>
<td>88.10</td>
<td>88.70</td>
<td>90.90</td>
<td>89.70</td>
<td>89.10</td>
</tr>
<tr>
<td>Angle $90^\circ$</td>
<td>94.00</td>
<td>93.70</td>
<td>96.00</td>
<td>96.10</td>
<td>94.10</td>
<td>94.40</td>
</tr>
<tr>
<td>Angle $135^\circ$</td>
<td>90.50</td>
<td>89.90</td>
<td>90.40</td>
<td>90.80</td>
<td>89.80</td>
<td>90.30</td>
</tr>
</tbody>
</table>

Figure 5.9: Classification accuracy of Tsallis Entropy at different values of $\alpha$
It is observed from Table 5.8 and Figure 5.9 that when $\alpha = 4$, Tsallis entropy produces better results at Angle $45^\circ$, Angle $90^\circ$, Angle $135^\circ$ and better results for Angle $0^\circ$ is produced at $\alpha = 2$. In Angle $0^\circ$ the classification accuracy at $\alpha = 2$ is 90.20% and when $\alpha = 4$, it is 87.40%. Hence, it is observed that the Tsallis Entropy produces better results at $\alpha = 4$.

The classification accuracy of Havrda and Charavat Entropy and the Tsallis Entropy is presented in Table 5.9 and is plotted in Figure 5.10.

### Table 5.9: Classification accuracy of Havrda and Charavat Entropy vs. Tsallis Entropy

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Havrda and Charavat Entropy</th>
<th>Tsallis Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angle 0°</td>
<td>86.40</td>
<td>89.20</td>
</tr>
<tr>
<td>Angle 45°</td>
<td>88.30</td>
<td>90.90</td>
</tr>
<tr>
<td>Angle 90°</td>
<td>95.80</td>
<td>96.10</td>
</tr>
<tr>
<td>Angle 135°</td>
<td>90.00</td>
<td>90.80</td>
</tr>
</tbody>
</table>

![Figure 5.10: Classification accuracy of Havrda and Charavat Entropy vs. Tsallis Entropy](image-url)
It is concluded from the Table 5.9 and Figure 5.9 that Tsallis Entropy produces better results comparing to Havrda and Charvat Entropy in all Angles.

The predictive accuracies of mammogram features at angles $0^\circ$, $45^\circ$, $90^\circ$ and $135^\circ$ using C4.5 and Ant-Miner and Tsallis Ant-Miner are listed in Table 5.10 and are plotted in Figure 5.11.

**Table 5.10: Report of Classification accuracy using C4.5, Ant-Miner and Ant-Miner-T**

<table>
<thead>
<tr>
<th>Data Set</th>
<th>C4.5 (%)</th>
<th>Ant-Miner (%)</th>
<th>Ant-Miner-T (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angle 0$^\circ$</td>
<td>90.80</td>
<td>88.60</td>
<td>89.20</td>
</tr>
<tr>
<td>Angle 45$^\circ$</td>
<td>90.30</td>
<td>90.40</td>
<td>90.90</td>
</tr>
<tr>
<td>Angle 90$^\circ$</td>
<td>95.20</td>
<td>94.80</td>
<td><strong>96.10</strong></td>
</tr>
<tr>
<td>Angle 135$^\circ$</td>
<td>88.80</td>
<td>90.10</td>
<td>90.80</td>
</tr>
</tbody>
</table>

**Figure 5.11: Classification accuracy using C4.5, Ant-Miner and Ant-Miner-T**
The results indicate that the Ant-Miner using Tsallis entropy achieves little higher accuracy rate in mammogram feature set at Angle $0^\circ$, Angle $45^\circ$ and Angle $90^\circ$ and Angle $135^\circ$.

The number of rules generated, TPR and FPR for mammogram features at angles $0^\circ$, $45^\circ$, $90^\circ$ and $135^\circ$ using C4.5 and Ant-Miner and Tsallis Ant-Miner are listed in Table 5.11.

Table 5.11: Number of Rules, TPR and FPR (Ant-Miner vs. Ant-Miner-T)

<table>
<thead>
<tr>
<th>Data Set</th>
<th>No. of Rules</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C4.5</td>
<td>AM</td>
<td>AM-T</td>
<td></td>
<td>TPR</td>
<td></td>
<td>FPR</td>
</tr>
<tr>
<td>Angle $0^\circ$</td>
<td>19.60</td>
<td>12.20</td>
<td>11.00</td>
<td></td>
<td>00.95</td>
<td>00.83</td>
<td>00.81</td>
</tr>
<tr>
<td>Angle $45^\circ$</td>
<td>21.00</td>
<td>10.00</td>
<td>09.90</td>
<td></td>
<td>00.94</td>
<td>00.94</td>
<td>00.95</td>
</tr>
<tr>
<td>Angle $90^\circ$</td>
<td>19.30</td>
<td>12.60</td>
<td>12.10</td>
<td></td>
<td>00.89</td>
<td>00.92</td>
<td>00.94</td>
</tr>
<tr>
<td>Angle $135^\circ$</td>
<td>10.30</td>
<td>12.50</td>
<td>11.30</td>
<td></td>
<td>00.83</td>
<td>00.85</td>
<td>00.86</td>
</tr>
</tbody>
</table>

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 $45^\circ$, Angle $90^\circ$ and Angle $135^\circ$. In all angles the false positive rate becomes better when compared to other classifiers.

As quoted by N. Rathie, Sergio Da Silva [138], the Tsallis nonextensive entropy of the statistical physics literature exactly matches the previously defined Havrda-Charvat structural $\alpha$-entropy of information theory, Tsallis entropy based Ant-Miner reports the approximately the same accuracy as reported by Kavrda and Charvat entropy. Since in the literature there is evidence for
performance improvement using Tsallis entropy this thesis proposes Ant-Miner-T which uses Tsallis entropy as its heuristic function.

5.9 SUMMARY

This chapter analysed the performance of different non-shannon entropy measures as the heuristic function of Ant-Miner. Different values of the parameter $\alpha$ 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 $0^\circ$, Angle $45^\circ$, Angle $90^\circ$, and Angle $135^\circ$ 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 $0^\circ$ 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 logarithmic calculation as it is needed in Shannon Entropy.