International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Unveiling the Secrets of the Mammogram: A Statistical Journey Through Machine Learning Approach

Dr. A. Jagadish Kumar[1], Prof. K. Karteeka Pavan[2], Prof (Retd.). A.V. Dattatreya Rao[3]

[1]Biostatistician, Dept. of Biostatistics & Statistical Programming, Sanofi Health Care India Private Limited, Hyderabad, Telangana, India
[2]Professor & Head, Dept. of Computer Applications, RVR&JC college of Engineering & Technology, Guntur, Andhra Pradesh, India
[3]Retired Professor of Statistics, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India

Abstract:
Mammograms, once mere X-ray images, are now powerful allies in the fight against breast cancer. Armed with sophisticated statistical image classification techniques, these images reveal hidden clues that can lead to early detection. Beneath the surface of a mammogram lies a complex tapestry of patterns, each one a potential marker of malignancy. By decoding these intricate patterns, algorithms can identify abnormalities that might otherwise go unnoticed. The Mammographic Image Analysis Society (MIAS) has been at the forefront of this revolution, creating a vast database of digital mammograms. From this treasure trove, we've selected 322 images to put our statistical prowess to the test. Using a powerful machine learning tool known as Support Vector Machine (SVM), we've classified these images into seven categories:

- **Calcifications (CALC):** Tiny, hard deposits.
- **Circumscribed (CIRC) masses:** Well-defined lumps.
- **Speculated (SPIC) masses:** Lumps with spiky edges.
- **Ill-defined (MISC) masses:** Irregular, indistinct lumps.
- **Architectural distortions (ARCH):** Changes in the breast's structure.
- **Asymmetry(s) (ASYM):** Differences between the two breasts.
- **Normal (NORM):** Healthy breast tissue.

Our goal is to shed light on the intricate process of mammogram classification and share the insights gained from our analysis presented in the user-friendly language and the analysis is carried out using MATLAB.

Keywords: Mammogram, GLCM, GLAM

I. Introduction

Broad steps involved in statistical analysis of mammogram are shown in the following block diagram.

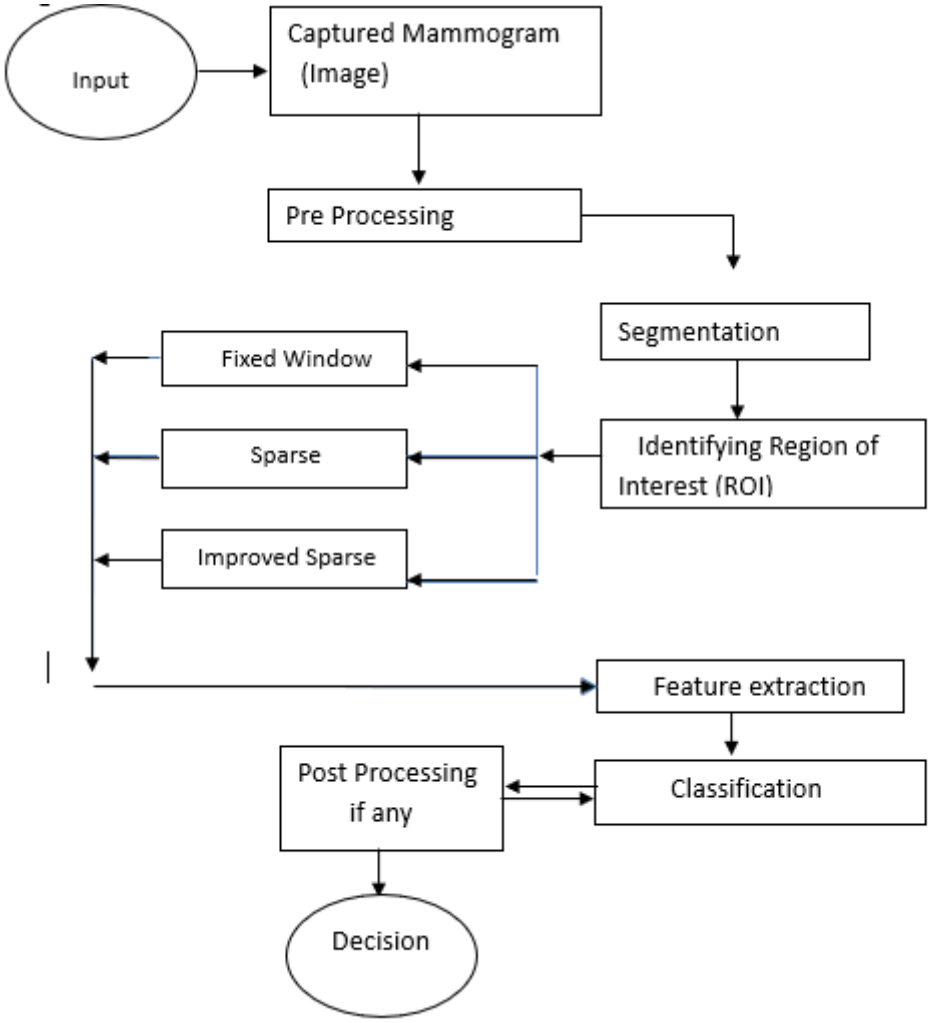

Figure 1: Block diagram of Mammogram Analysis from Pattern Recognition point of view Details of preprocessing, identification of Region of Interest (ROI), feature extraction and classification are presented in subsequent sections from 2 to 5.

II. Preprocessing of Mammogram Images

To enhance the diagnostic accuracy of mammograms, **pre-processing** is essential. This involves removing various **artifacts** and **noise** that can obscure critical features. Specific pre-processing techniques include:
Digitization noise reduction: Removing errors introduced during the image acquisition process.
Artifact suppression: Eliminating spurious structures or patterns that are not related to breast tissue.
Background separation: Isolating the region of interest (breast tissue) from the surrounding background.
Pectoral muscle segmentation: Identifying and separating the pectoral muscle from the breast tissue.
These steps collectively improve the quality of the mammogram image, making it easier for radiologists to detect abnormalities.

III. Finding Region of Interest (ROI)

A. Using Image Segmentation

[Qinghua Huang, et.al., 2017] define image segmentation as the process of dividing an image into regions to separate objects from the background. There are four main methods for image segmentation:

Thresholding: This method separates an image into different regions based on a predetermined gray level threshold. Pixels with values above the threshold are considered part of the object, while those below are considered background.

Edge Detection: This method focuses on identifying boundaries between objects. By analyzing discontinuities in gray level values, edges and contours within the image are detected. These edges play a crucial role in extracting features for object recognition.

Region-Based Segmentation: This method groups pixels with similar gray levels into distinct regions. Pixels that are connected and share similar intensity values are grouped together.

Region Growing: This method considers the similarity of neighboring pixels. Pixels with common characteristics are grouped together to form regions with a certain degree of homogeneity.

The underlying principle of region segmentation is to subdivide an image into smaller, non-overlapping areas with consistent properties. Figure 1 illustrates the spatial decomposition of an image into four separate regions. The union of these regions must encompass the entire image domain, while their intersection should be empty (no pixels belong to both regions simultaneously).

Formalizing Segmentation: Let R_T represent the entire image region. We can express image segmentation as a process that partitions R_T into n non-overlapping sub-regions, denoted as R_1, R_2, ..., R_i, ..., R_n (refer to Figure 1). Here are the key mathematical properties of this segmentation:

i) **Completeness:** Every pixel in the image must belong to one and only one region. Mathematically, this translates to:

$$R_T = \bigcup_{i=1}^{n} R_i$$ (Union of all Ri regions is equal to the entire image region)

ii) **Connectedness:** Pixels within a region must be connected to each other.

iii) **Disjointness:** Regions must not overlap. No pixel can belong to both R_i and R_j simultaneously, where $i \neq j$. Mathematically:

$R_i \cap R_j = \emptyset$ (Intersection of any two regions is the null set) for $i \neq j$

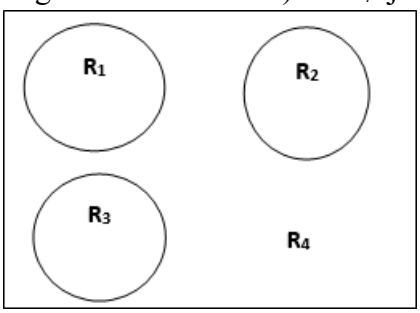

Figure 2: The Spatial decomposition of an image into smaller regions

iv) **Regional Properties:** This condition refers to specific characteristics that pixels within a region must satisfy. For example, all pixels in R_i might have the same intensity value.

v) **Mutual Exclusion:** No two regions can share the same properties. Regions R_i and R_j must be distinct based on the defined predicate P. Mathematically:

$P(R_i)$ = TRUE for i = 1, 2, ..., n (Predicate P holds true for all pixels in R_i)

$P(R_i \cup R_j)$ = FALSE for $i \neq j$ (Predicate P is false for the union of R_i and R_j)

where $P(R_i)$ is a logical predicate applied to the pixels in set Ri and Ø represents the null set.

Explanation:

Condition (i) emphasizes the requirement for a complete segmentation, meaning every pixel needs to be

assigned to a specific region.
Condition (ii) ensures that pixels within a region are spatially connected.
Condition (iii) guarantees that regions are distinct and do not overlap.
Condition (iv) defines the properties that pixels within a region should exhibit.
Condition (v) underlines the concept of mutual exclusion, where no two regions can share the same characteristics based on the defined predicate.

B. Using Window Texture Analysis

From the segmented image, the Region of Interest (ROI) is identified as the cluster containing the brightest pixels. To extract features from this ROI, a traditional approach involves selecting a fixed-size window that encompasses the ROI. This window size is determined automatically based on the ROI's dimensions. To accelerate processing, ROIs were extracted from the MIAS database. To comply with radiologist recommendations, these ROIs were resized to a maximum of 64x64 pixels, with original sizes varying from 20x20 pixels. The minimum and maximum x and y coordinates of the ROI(s) are used to calculate the window size, resulting in a 64x64 pixel window.

C. Sparse Texture Analysis

In computer-aided diagnosis (CAD) systems, regions of interest (ROIs) are typically identified using image segmentation techniques (Chu et al., 2015). Due to the diverse sizes and shapes of abnormalities, segmented ROIs often vary in dimensions. Traditionally, researchers have resized each ROI into a fixed-size window for feature extraction. However, this approach can be suboptimal as mass regions are inherently irregular and variable in size.

The high dimensionality of the feature space resulting from different-sized ROIs can degrade classification performance. Using a large common template for ROIs not only lowers accuracy but also increases computational cost (Antonio et al., 2013). Researchers often select different window sizes based on factors such as the study type, dataset, and malignancy class of the mammogram. Determining the optimal ROI size for a specific class typically involves experimentation with various window sizes.

To address these limitations, this research proposes a novel method for representing arbitrary-shaped ROIs automatically: Sparse ROIs. This approach focuses solely on the mass region, excluding non-contributing pixels from the segmentation process. By eliminating unnecessary information, Sparse ROIs significantly reduce the feature space and computational time.

This paper also discusses the importance of optimal ROI size, the ability to locate irregular-shaped ROIs using sparse techniques, and effective feature extraction methods.

The cluster from the earlier step of selected ROI is modeled using sparse representation to cover the irregular shaped mass as shown in Figure 3.

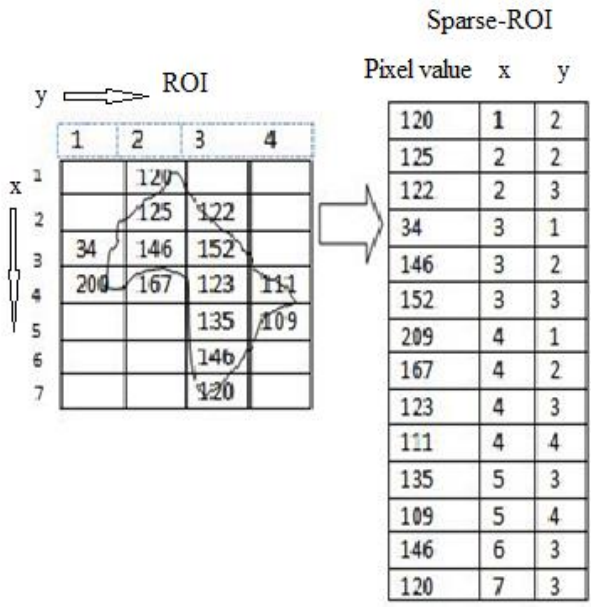

Figure 3: Sparse Representation

D. Improved Sparse texture Analysis

The Sparse ROI, introduced in 3.3, is a vectorized representation of the original two-dimensional ROI. Factors such as the range (number of gray levels), n (number of pixels) and column width (maximum width or maximum number of columns of mass (ROI)) can significantly impact processing time. To address this, we propose a new enhanced Sparse ROI model that completely eliminates the influence of column width, resulting in reduced computational complexity. Unlike the Sparse ROI model, the improved sparse model retains all four neighboring pixels and the current pixel within the same row, instead of using x and y coordinates (as shown in Figure 3). Figure 4 demonstrates how this improved Sparse ROI method can effectively represent irregularly shaped masses (clusters of bright spots).

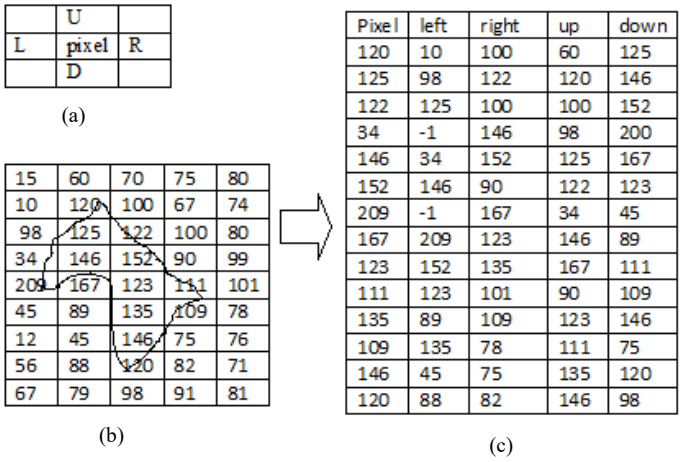

Figure 4: Improved Sparse Representation

(a) Pixel and its neighbors L (left), R (right), U (up) and D (down).
(b) Sample ROI
(c) Improved Sparse-ROI

III. Construction of the Gray-Level Co-occurrence Matrix (GLCM) and Gray-Level Aura Matrix (GLAM)

This section describes a method for analyzing textures in images, called Statistical Texture Analysis. Here's a breakdown of flow of evaluating GLCM:

a) Statistical Texture Analysis: Statistical Texture Analysis extract features that describe the texture of an image. It focuses on the statistical distribution of how pixel intensities are arranged in specific locations relative to each other. The number of intensity levels (pixels) in each combination influences the process.

b) Gray-Level Co-occurrence Matrix (GLCM): The GLCM is a popular technique for extracting second-order statistical texture features (Kanchan Lata Kashyap et al., 2017). This matrix has rows and columns corresponding to the number of gray levels in the image. Considering the following image of size 5x5 with 4 gray levels is,

$$\begin{bmatrix} 0 & 1 & 1 & 2 & 3 \\ 0 & 0 & 2 & 3 & 3 \\ 0 & 1 & 2 & 2 & 3 \\ 1 & 2 & 3 & 2 & 2 \\ 2 & 2 & 3 & 3 & 2 \end{bmatrix}$$

c) Normalized GLCM and Relative Frequencies: When normalized, each element in the GLCM matrix, P (i, j/Δx, Δy), represents the relative frequency of two pixels with intensities 'i' and 'j' appearing next to each other, separated by a distance of (Δx, Δy). In simpler terms, it captures how often a specific combination of gray levels occurs within a certain distance (d)and direction (θ) in the image.

d) Calculating GLCM Elements: Given an image neighborhood (a small region) containing 'L' gray levels (from 0 to L-1), and f(m, n) representing the intensity at position (m, n) within the neighborhood, the formula for P(i, j/Δx, Δy) is: P(i, j/Δx ,Δy) = W * Q(i, j/Δx, Δy) where W is a normalization factor and Q(i, j/Δx, Δy) represents the number of times the specific pixel combination (i, j) occurs at the given distance (Δx, Δy).

e) Calculating GLCM for a 5x5 Image: For the above 5x5 image by using the definition of GLCM and considering the relative frequencies P (i, j ; 1, 0) for i = 0 (1) 3 and j= 0 (1) 3, the 4x4 GLCM is as under (since we have taken 4 gray levels the size of GLCM is 4).

P (i, j ; 1, 0)

i \ j	0	1	2	3
0	1/20	2/20	1/20	0
1	0	1/20	3/20	0
2	0	0	3/20	5/20
3	0	0	2/20	2/20

f) Challenges of GLCM and Reducing Gray Levels: As the number of gray levels increases, the GLCM matrix size grows significantly, requiring more storage space. This can lead to situations where the feature matrix is larger than the original image! Additionally, GLCMs are sensitive to the size of

the analyzed image region. To address these limitations, researchers often choose to reduce the number of gray levels used in the analysis.

g) **GLCM Parameterization and Angular Relationships:** GLCM analysis requires specifying parameters such as the distance (d) and angle (θ) between the pixels being compared. Due to the repetitive nature of textures, often only a limited number of angles are considered (typically 0°, 45°, 90°, and 135°). Figure 5 illustrates the geometrical relationships between these parameters.

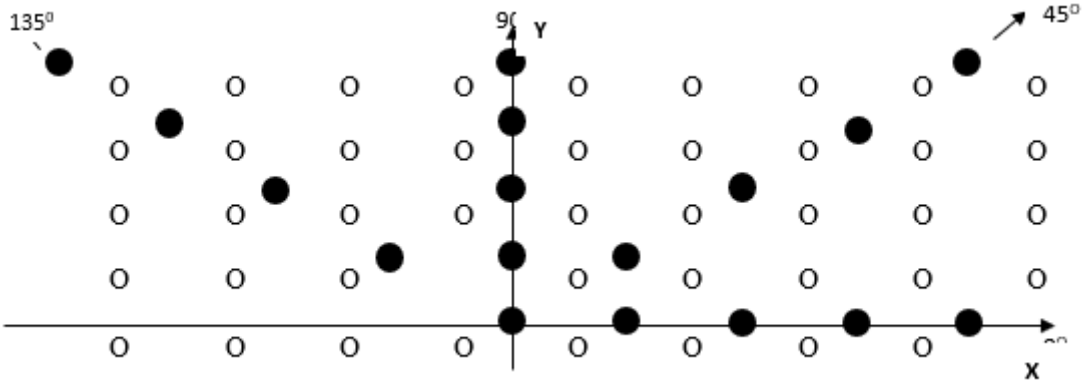

Figure 5: Geometrical relationship of GLCM

To obtain a statistically reliable estimate of the joint probability distribution, the co-occurrence matrix must contain a sufficiently high average occupancy level. This can be achieved by either reducing the number of gray level quantization levels or by using a relatively large window size.

Reducing the number of gray levels can compromise the accuracy of texture description, especially for low-amplitude textures. Conversely, using a large window size can introduce uncertainty and error, particularly in regions with significant texture variations.

A common approach to balance these trade-offs is to use 16 gray levels and a window size of approximately 32 to 64 pixels on each side.

Simple relationships exist among certain pairs of the estimated probability distributions P (d,θ). Let $P^T(d,\theta)$ denote the transpose of the matrix P (d,θ). Then

$$P(d,0^0) = P^T(d,180^0)$$
$$P(d,45^0) = P^T(d,225^0)$$
$$P(d,90^0) = P^T(d,270^0)$$
$$P(d,135^0) = P^T(d,315^0)$$

Thus, the knowledge of P $(d, 180^0)$, P $(d, 225^0)$, P $(d, 270^0)$ and P $(d, 315^0)$ adds nothing to the specification of the texture. In general, if $X \in Z^n$, the number of independent directions from X in Z^n is $\frac{3^n - 1}{2}$. For n=2, number of independent directions is equal to four. Hence, we have taken 0^0, 45^0, 90^0, and 135^0.

For n=2, number of independent directions is equal to four. Hence, we have taken 0^0, 45^0, 90^0, and 135^0. Then GLAM can be represented as sum of GLCMs of the said 4 directions for fixed distance.

IV. Textural Feature Extraction from GLCM or GLAM

We use the following notations in describing features (Zyout Imad, et. al., 2015) from GLCM.
Let us represent matrix P_{ij} as follows

i \ j	1	2	…	K
1	P_{11}	P_{12}	…	P_{1k}
2	P_{12}	P_{22}	…	P_{2k}
. . .			. . .	
k	P_{k1}	P_{k2}	...	P_{kK}

Row wise marginal totals are represented as $P_{1.}$, $P_{2.}$, ..., $P_{k.}$ and those of column wise are $P_{.1}$, $P_{.2}$, ..., $P_{.K}$ such that

$$\sum_{i=1}^{k} p_{i.} = \sum_{j=1}^{K} p_{.j} = p_{..}$$

$$\mu_x = E(\mathrm{i}) = \sum_{i=1}^{k} i p_{i.} = \text{marginal mean w.r.t 'x'}$$

$$\mu_y = E(j) = \sum_{j=1}^{K} j p_{.j} = \text{marginal mean w.r.t 'y'}$$

μ =over all mean

$$\sigma_x^2 = \sum_{i=1}^{k} (i-\mu_x)^2 p_{i.} \quad , \quad \sigma_y^2 = \sum_{j=1}^{K} (j-\mu_y)^2 p_{.j}$$

$$\sigma_{xy} = \sum_{i=1}^{k} \sum_{j=1}^{K} (i-\mu_x)(j-\mu_y) p_{ij}$$

The following textural features are extracted.

i) Angular Second Moment (ASM) = $\sum_i \sum_j P_{ij}^2$

ASM is a measure of image homogeneity. A homogeneous image contains few gray levels, resulting in a GLCM with a few but relatively high P (i, j) values. This leads to a high sum of squares.

ii) Contrast = $\sum_{n=0}^{L-1} n^2 \left(\sum_{i=1}^{L} \sum_{j=1}^{L} P_{ij} \right)$, $|i-j| = n$

This measure quantifies the contrast or local intensity variation, which is primarily influenced by the off-diagonal elements of the GLCM.

iii) Correlation = $\frac{\sigma_{xy}}{\sigma_x.\sigma_y}$

Correlation measures the linear relationship between the gray levels of pixels at specified relative positions within an image.

iv) Total Variance $(\sigma^2) = \sum_{i=1}^{k}\sum_{j=1}^{k}(i-\mu)^2 P_{ij}$

This feature emphasizes the elements that deviate significantly from the average value of P_{ij}.

v) Inverse Difference Moment (IDM)

$$\text{IDM}= \sum_{i=1}^{k}\sum_{j=1}^{k}\frac{1}{1+(i-j)^2}P_{ij}$$

IDM is also influenced by image homogeneity. The $(1+(i-j)^2)^{-1}$ weighting scheme assigns lower weights to off-diagonal elements. As a result, IDM tends to be lower for heterogeneous images and higher for homogeneous images.

vi) Sum Average

$$\text{AVER} = \sum_{i=0}^{2L-2} iP_{x+y}(i)$$

The Sum Average, like variance, is rotationally invariant.

vii) Entropy

$$\text{ENTROPY} = -\sum_{i=0}^{L-1}\sum_{j=0}^{L-1}P_{ij}.\log P_{ij}$$

Heterogeneous scenes typically have low first-order entropy, while homogeneous scenes tend to have higher entropy.

viii) Sum Entropy (SENT)

$$\text{SENT}= -\sum_{i=0}^{2L-2} P_{x+y}(i)\log(P_{(x+y)}(i))$$

Sum Entropy is sensitive to image rotation and can be used to differentiate between images with different rotational orientations.

ix) Sum Variance (SV)

$$\text{SV}= \sum_{i=0}^{2L-2}(i-SENT)^2 P_{(x+y)}(i)$$

It measures the dispersion of gray-level values. A higher Sum Variance indicates a greater spread of gray-level values in the image, suggesting a more heterogeneous texture. Conversely, a lower Sum Variance implies a more homogeneous texture with less variation in gray levels.

x) Difference Entropy

$$\text{DENT}= -\sum_{i=0}^{L-1} P_{x+y}(i)\log(P_{x+y}(i))$$

Difference Entropy is a statistical texture feature derived from the Gray-Level Co-occurrence Matrix (GLCM). It measures the randomness or disorder of the gray-level differences between neighboring pixels.

A higher Difference Entropy indicates a greater degree of randomness or variation in the differences between neighboring pixels, suggesting a more complex texture. Conversely, a lower Difference Entropy implies a more ordered and uniform texture.

xi) Inertia= $\sum_{i=0}^{L-1}\sum_{j=0}^{L-1}(i-j)^2 P_{ij}$

Inertia is a statistical texture feature derived from the Gray-Level Co-occurrence Matrix (GLCM). It measures the concentration of elements around the diagonal of the GLCM. A higher Inertia value indicates that the gray-level values of neighboring pixels are more similar, suggesting a more homogeneous texture. Conversely, a lower Inertia value implies a more heterogeneous texture with greater variation in gray-level differences between neighboring pixels.

xii) Cluster Shade

$$\text{SHADE}= \sum_{i=0}^{L-1}\sum_{j=0}^{L-1}(i+j-\mu_x-\mu_y)^3 P_{ij}$$

Cluster Shade is a statistical texture feature derived from the Gray-Level Co-occurrence Matrix (GLCM). It measures the skewness of the GLCM, indicating the asymmetry of the distribution of gray-level pairs. A higher Cluster Shade value suggests a more skewed distribution, indicating a texture with a dominant direction or pattern. A lower Cluster Shade value, on the other hand, indicates a more symmetric distribution, suggesting a more uniform texture.

xiii) Cluster Prominence

$$\text{PROM}= \sum_{i=0}^{L-1}\sum_{j=0}^{L-1}(i+j-\mu_x-\mu_y)^4 P_{ij}$$

Cluster Prominence is a statistical texture feature derived from the Gray-Level Co-occurrence Matrix (GLCM). It measures the strength and number of clusters of high probability values within the GLCM. A higher Cluster Prominence indicates a more pronounced clustering of gray-level values, suggesting a more regular and repetitive texture. Conversely, a lower Cluster Prominence indicates a more random and less structured texture.

V. Classification of Mammograms and performance measures

Mammogram Classification using Multiclass SVM: A multiclass Support Vector Machine (SVM) was employed to classify mammograms into seven categories based on thirteen extracted features. To assess the performance of various feature extraction schemes, a confusion matrix was generated for each experiment with 90%, 80%, 70%, and 60% of the data used for training, with the remainder used for testing under various schemes mentioned below.

i)Fixed Window: Gray-Level Co-occurrence Matrix (GLCM) and Gray-Level Aura Matrix (GLAM) were extracted from fixed-size windows within each mammogram.

ii)Sparse Window: GLCM and GLAM were extracted from sparse windows within each mammogram.

iii)Improved Sparse Window: GLAM was extracted from improved sparse windows within each mammogram.

Dataset: The dataset comprised 322 mammogram images, distributed unevenly across seven classes. For each classification experiment, a portion (p%) of images was randomly selected from each class for

training, while the remaining (1-p) % served as the test set.

Performance Metrics: The following performance metrics were calculated based on the confusion matrix:

Accuracy: Overall correct classification rate

Sensitivity: True Positive Rate (TPR)

Specificity: True Negative Rate (TNR)

Precision: Positive Predictive Value (PPV)

VI. Results

A fixed window size of 64x64 pixels was used to segment Regions of Interest (ROIs) in all 322 mammograms. This approach yielded the following results:

Pixel Count: Number of pixels within the fixed window ROI used for constructing Gray-Level Co-occurrence Matrices (GLCM) and Gray-Level Aura Matrices (GLAM), as well as the number of pixels in Sparse and Improved Sparse methods.

Computational Time: Time taken by traditional Fixed Window, Sparse, and Improved Sparse ROI methods for GLCM and GLAM construction.

Extracted Textural Features: Textural features extracted using Fixed Window GLCM, Fixed Window GLAM, Sparse GLCM, Sparse GLAM, and Improved Sparse GLAM.

To conserve space, only the results for 25 mammograms, including pixel counts, computation time, and extracted textural features, are presented here in Tables 1-9. The complete results are available upon request from the first author.

Table1: No. of pixels in Fixed Window ROI for the construction of GLCM and GLAM

Sl. No.	Fixed Window		
	No. of Rows	No. of Columns	No. of Pixels
1	20	51	1020
2	421	435	183135
3	113	211	23843
4	568	523	297064
5	339	175	59325
6	29	36	1044
7	6	5	30
8	506	6	3036
9	238	262	62356
10	492	348	171216
11	42	61	2562
12	635	274	173990
13	412	286	117832
14	269	468	125892
15	297	341	101277
16	349	201	70149
17	310	108	33480

18	470	310	145700
19	715	448	320320
20	230	299	68770
21	688	390	268320
22	332	394	130808
23	449	346	155354
24	631	242	152702
25	500	271	135500

Table 2: Number of pixels in Sparse/Improved Sparse methods

S. No.	**Sparse/Improved Sparse**
	No. of pixels
1	67661
2	82628
3	113343
4	87241
5	171
6	258
7	2144
8	3060
9	23815
10	839
11	24723
12	72
13	10885
14	239099
15	159567
16	17037
17	117
18	15005
19	4366
20	3405
21	699
22	15894
23	3736
24	1718
25	120820

Table3: Time taken by traditional Fixed Window, Sparse and Improved Sparse ROIs for GLCM and GLAM

S. No.	GLCM (time in sec.)		GLAM (time in sec.)		
	Fixed window	Sparse	Fixed window	Sparse	Improved Sparse
1	143.5852	2.957019	260.482	1352.123	2.955628
2	145.2275	4.448406	268.2461	1882.818	3.904705
3	164.9035	6.460743	298.6097	2054.91	5.458111
4	188.369	2.515627	302.0604	1113.971	1.799991
5	191.4232	0.023025	423.9706	0.044355	0.005023
6	220.9597	0.022237	406.6209	0.225772	0.009506
7	205.5852	0.148843	314.8274	5.053507	0.024971
8	275.5019	0.586013	387.4853	34.73158	0.131029
9	259.5557	0.319164	305.5605	357.4795	0.45015
10	271.5191	0.036462	243.8996	0.369994	0.011707
11	262.0639	0.069719	315.1551	385.7937	0.534382
12	271.3971	0.04036	295.1555	0.008339	0.004593
13	262.7504	0.205182	395.8247	65.91151	0.129517
14	359.9116	12.99278	440.8307	7232.642	12.87329
15	352.7992	6.321749	269.7271	3854.157	6.409962
16	348.723	0.451405	235.8287	187.4673	0.315695
17	340.0404	0.009062	154.3511	0.037845	0.005808
18	347.0623	0.574854	157.4065	154.0435	0.437949
19	346.0856	0.07669	457.7204	14.92219	0.027928
20	342.2221	0.119756	403.0937	19.70896	0.054855
21	341.8415	0.377752	286.6165	1.420357	0.035128
22	340.6856	0.339082	258.2849	97.33222	0.250247
23	339.8158	0.071054	341.4162	5.57726	0.033975
24	339.6045	0.295753	337.6345	26.3731	0.281756
25	338.9166	5.626249	449.4328	4183.488	6.043973

Table 4: Textural Features Extracted by Fixed Window GLCM for 25 images

S. No.	ASM	CONTR-AST	IDM	CORREL-ATION	DISSIMI-LARITY	ENTROPY	SUM OF SQUARES OF VARIANCE	INERTIA	CLUSTER SHADE	CLUSTER PROMINENCE	DIFFERENCE ENTROPY	SUM ENTROPY	SUM AVERAGE
1	9.50E-04	5.60E+00	4.73E-01	9.99E-01	1.64E+00	3.10E-03	2.13E+04	5.60E+00	2.87E+07	1.01E+10	2.74E-02	3.80E-02	1.36E+02
2	2.02E-03	4.26E+00	5.47E-01	9.99E-01	1.34E+00	-2.21E-02	3.08E+04	4.26E+00	4.66E+07	1.81E+10	1.56E-02	1.73E-01	1.69E+02
3	2.65E-03	1.21E+01	5.48E-01	9.98E-01	1.46E+00	8.07E-03	3.53E+04	1.21E+01	5.67E+07	2.31E+10	-3.41E-03	-1.14E-01	1.81E+02
4	2.05E-03	3.40E+00	5.59E-01	9.99E-01	1.21E+00	8.06E-03	3.10E+04	3.40E+00	4.68E+07	1.81E+10	8.83E-03	7.70E-02	1.70E+02
5	2.05E-03	3.42E+00	5.59E-01	9.99E-01	1.21E+00	1.08E-02	3.10E+04	3.42E+00	4.67E+07	1.80E+10	8.83E-03	7.90E-02	1.70E+02
6	4.79E-03	2.78E+00	5.80E-01	9.99E-01	1.11E+00	-1.78E-02	2.55E+04	2.78E+00	3.45E+07	1.21E+10	9.32E-03	-8.59E-02	1.56E+02
7	3.67E-03	3.31E+00	5.63E-01	9.99E-01	1.19E+00	1.93E-04	2.61E+04	3.31E+00	3.58E+07	1.26E+10	5.12E-03	-8.81E-02	1.58E+02
8	2.10E-03	5.04E+00	4.77E-01	9.98E-01	1.56E+00	1.94E-02	2.22E+04	5.04E+00	2.82E+07	9.17E+09	-4.61E-02	3.75E-02	1.44E+02
9	1.76E-03	7.61E+00	4.85E-01	9.98E-01	1.56E+00	-1.28E-02	2.12E+04	7.61E+00	2.67E+07	8.68E+09	3.74E-02	-3.73E-02	1.39E+02
10	1.74E-03	7.70E+00	4.85E-01	9.98E-01	1.57E+00	1.60E-02	2.12E+04	7.70E+00	2.68E+07	8.72E+09	3.74E-02	7.35E-02	1.39E+02
11	1.74E-03	7.71E+00	4.85E-01	9.98E-01	1.57E+00	-1.07E-03	2.12E+04	7.71E+00	2.68E+07	8.73E+09	3.74E-02	-5.41E-02	1.39E+02

12	1.73E-03	1.19E+01	4.85E-01	9.97E-01	1.62E+00	-4.38E-04	2.13E+04	1.19E+01	2.70E+07	8.79E+09	-1.12E-03	3.15E-02	1.40E+02
13	1.21E-03	1.39E+01	4.89E-01	9.97E-01	1.64E+00	-1.47E-02	2.40E+04	1.39E+01	3.27E+07	1.15E+10	-6.71E-03	2.36E-02	1.48E+02
14	2.30E-03	1.60E+01	5.94E-01	9.98E-01	1.20E+00	-1.21E-02	2.46E+04	1.60E+01	3.46E+07	1.25E+10	5.48E-02	6.42E-02	1.46E+02
15	1.49E-03	4.37E+01	4.77E-01	9.91E-01	1.85E+00	4.27E-02	2.29E+04	4.37E+01	3.04E+07	1.03E+10	-9.41E-02	1.91E-01	1.43E+02
16	1.47E-03	5.04E+01	4.76E-01	9.87E-01	1.90E+00	7.91E-03	2.28E+04	5.04E+01	2.97E+07	9.90E+09	-4.46E-02	6.74E-02	1.45E+02
17	1.46E-03	5.10E+01	4.77E-01	9.87E-01	1.91E+00	-2.51E-02	2.30E+04	5.10E+01	3.00E+07	1.00E+10	-4.49E-02	-9.48E-03	1.45E+02
18	2.52E-03	8.17E+01	4.97E-01	9.89E-01	2.08E+00	2.50E-02	2.09E+04	8.17E+01	2.78E+07	9.51E+09	-2.39E-01	2.62E-01	1.32E+02
19	2.55E-03	6.26E+01	5.16E-01	9.90E-01	1.86E+00	-4.60E-02	2.28E+04	6.26E+01	3.07E+07	1.06E+10	2.41E-01	1.67E-01	1.40E+02
20	2.70E-03	3.21E+01	5.20E-01	9.95E-01	1.66E+00	5.55E-02	2.20E+04	3.21E+01	2.91E+07	9.79E+09	1.37E-01	1.54E-01	1.38E+02
21	2.27E-03	3.31E+01	4.96E-01	9.97E-01	1.78E+00	-5.56E-02	2.16E+04	3.31E+01	2.86E+07	9.67E+09	-1.63E-01	-1.97E-01	1.36E+02
22	2.15E-03	3.50E+01	4.99E-01	9.95E-01	1.80E+00	-4.00E-02	2.25E+04	3.50E+01	3.04E+07	1.05E+10	1.67E-01	-2.38E-01	1.39E+02
23	2.11E-03	3.52E+01	5.01E-01	9.95E-01	1.80E+00	-4.10E-02	2.26E+04	3.52E+01	3.08E+07	1.08E+10	1.67E-01	-4.83E-02	1.39E+02
24	2.29E-03	3.53E+01	5.37E-01	9.99E-01	1.67E+00	-3.82E-03	1.99E+04	3.53E+01	2.61E+07	8.79E+09	-1.88E-01	2.29E-01	1.28E+02
25	1.59E-03	2.21E+01	5.31E-01	9.99E-01	1.43E+00	-1.54E-02	1.85E+04	2.21E+01	2.22E+07	7.04E+09	2.50E-02	2.57E-02	1.30E+02

Table5: Textural Features of 25 images extracted using Fixed Window GLAM

S.No	ASM	CONTR-AST	IDM	CORREL-ATION	DISSIMI-LARITY	ENTROPY	SUMOFSQUARESOFVARIANCE	INERTIA	CLUSTERSHADE	CLUSTERPROMINENCE	DIFFERENCEENTROPY	SUMENTROPY	SUMAVERAGE
1	4.65E+00	5.39E-03	5.46E-01	1.00E+00	1.39E+00	1.32E-01	1.50E+04	4.65E+00	1.97E+07	6.86E+09	-2.55E-01	2.35E-01	1.01E+02
2	4.47E+00	1.65E-03	5.62E-01	1.00E+00	1.30E+00	-1.11E-01	2.00E+04	4.47E+00	2.89E+07	1.10E+10	-9.13E-02	-1.42E-01	1.22E+02
3	1.47E+01	3.88E-03	5.67E-01	9.99E-01	1.41E+00	-2.75E-01	2.25E+04	1.47E+01	3.38E+07	1.32E+10	-4.01E-01	-5.04E-01	1.28E+02
4	1.97E+01	2.37E-03	5.55E-01	9.98E-01	1.42E+00	1.06E-01	2.49E+04	1.97E+01	3.68E+07	1.40E+10	-1.79E-01	1.75E-01	1.42E+02
5	1.10E+01	2.95E-03	6.02E-01	9.98E-01	1.14E+00	2.27E-02	1.37E+04	1.10E+01	1.51E+07	4.27E+09	-1.80E-01	1.07E-01	1.05E+02
6	1.25E+01	4.36E-03	5.57E-01	9.96E-01	1.27E+00	-3.49E-02	1.73E+04	1.25E+01	1.97E+07	5.73E+09	-6.62E-02	-9.65E-02	1.24E+02
7	1.34E+01	2.75E-03	5.24E-01	9.98E-01	1.48E+00	1.15E-01	1.63E+04	1.34E+01	1.97E+07	6.13E+09	-3.50E-01	3.66E-01	1.13E+02
8	1.60E+01	1.82E-03	4.83E-01	9.97E-01	1.68E+00	-8.81E-02	1.83E+04	1.60E+01	2.22E+07	6.89E+09	-9.73E-02	-1.20E-01	1.26E+02
9	1.60E+01	4.30E-03	5.68E-01	9.98E-01	1.33E+00	3.41E-01	1.56E+04	1.60E+01	1.87E+07	5.70E+09	-3.23E-01	3.52E-01	1.10E+02
10	3.85E+00	5.39E-03	5.38E-01	1.00E+00	1.33E+00	1.59E-01	1.75E+04	3.85E+00	2.21E+07	7.14E+09	-1.34E-01	1.19E-01	1.16E+02

11	3.44E+00	3.91E-03	5.53E-01	1.00E+00	1.24E+00	1.68E-01	1.90E+04	3.44E+00	2.45E+07	8.11E+09	-1.61E-01	2.48E-01	1.24E+02
12	2.83E+00	7.58E-03	5.89E-01	1.00E+00	1.11E+00	-1.60E-01	1.47E+04	2.83E+00	1.72E+07	5.12E+09	-4.26E-01	-3.66E-01	1.06E+02
13	5.69E+00	1.60E-03	5.59E-01	9.99E-01	1.25E+00	-3.60E-02	2.13E+04	5.69E+00	2.92E+07	1.03E+10	-7.15E-02	-1.00E-01	1.32E+02
14	2.32E+00	2.50E-03	6.12E-01	1.00E+00	1.00E+00	1.29E-01	2.15E+04	2.32E+00	3.01E+07	1.09E+10	-2.36E-01	3.72E-01	1.30E+02
15	5.71E+00	1.74E-03	4.63E-01	9.99E-01	1.67E+00	6.92E-02	2.17E+04	5.71E+00	2.85E+07	9.55E+09	-8.46E-02	5.57E-02	1.38E+02
16	5.22E+00	2.47E-03	4.67E-01	9.99E-01	1.63E+00	-1.27E-01	1.82E+04	5.22E+00	2.24E+07	7.06E+09	-1.35E-01	-1.42E-01	1.24E+02
17	4.99E+00	1.51E-02	5.38E-01	1.00E+00	1.43E+00	1.57E-01	1.93E+04	4.99E+00	2.65E+07	9.28E+09	-3.91E-01	3.81E-01	1.19E+02
18	4.34E+00	2.34E-02	5.84E-01	1.00E+00	1.27E+00	-5.92E-02	1.66E+04	4.34E+00	2.23E+07	7.62E+09	-5.54E-01	-5.40E-01	1.05E+02
19	7.30E+00	1.56E-03	5.02E-01	9.99E-01	1.47E+00	4.32E-02	2.14E+04	7.30E+00	2.79E+07	9.31E+09	-5.74E-03	1.00E-02	1.37E+02
20	4.07E+00	2.51E-03	5.11E-01	9.99E-01	1.40E+00	4.75E-02	1.99E+04	4.07E+00	2.51E+07	8.09E+09	-9.63E-02	1.15E-01	1.31E+02
21	4.86E+00	2.97E-03	5.26E-01	9.99E-01	1.43E+00	9.26E-02	1.82E+04	4.86E+00	2.37E+07	7.94E+09	-1.99E-01	2.33E-01	1.19E+02
22	3.64E+00	1.13E-02	5.70E-01	1.00E+00	1.25E+00	-1.09E-01	1.73E+04	3.64E+00	2.32E+07	7.98E+09	-4.95E-01	-5.16E-01	1.10E+02
23	2.50E+	3.96E	5.96E	1.00E+	1.06E	-3.23	2.24E+0	2.50E	3.11E+	1.10E+	4.12E-02	-4.20	1.35E+

	00	-03	-01	00	+00	E-01	4	+00	07	10		E-02	02
24	2.59E+00	4.85E-03	6.23E-01	1.00E+00	9.93E-01	-8.72E-02	1.73E+04	2.59E+00	2.25E+07	7.49E+09	1.20E-02	7.43E-02	1.13E+02
25	4.32E+00	1.69E-03	5.48E-01	9.99E-01	1.34E+00	6.08E-02	1.24E+04	4.32E+00	1.37E+07	4.05E+09	-1.59E-01	-1.44E-01	9.82E+01

Table 6: Textural Features Extracted by Sparse GLCM for 25 images

S.No.	ASM	CONTR-AST	IDM	CORREL-ATION	DISSIMI-LARITY	ENTROPY	SUM OF SQUARES OF VARIANCE	INERTIA	CLUSTER SHADE	CLUSTER PROMINENCE	DIFFERENCE ENTROPY	SUM ENTROPY	SUM AVERAGE
1	4.28E-03	2.90E+00	5.57E-01	9.93E-01	1.18E+00	-3.68E-02	8.44E+02	2.90E+00	2.51E+05	2.00E+07	-4.31E-02	-4.31E-02	2.55E+01
2	7.45E-03	1.68E+00	6.65E-01	9.96E-01	8.22E-01	2.26E-02	9.66E+02	1.68E+00	2.96E+05	2.39E+07	5.28E-02	5.28E-02	2.77E+01
3	6.07E-03	1.94E+00	6.34E-01	9.94E-01	9.12E-01	-3.29E-02	1.95E+03	1.94E+00	7.66E+05	7.86E+07	-3.62E-01	-3.62E-01	4.23E+01
4	1.19E-02	1.37E+00	6.71E-01	9.84E-01	7.67E-01	2.48E-02	8.33E+02	1.37E+00	2.07E+05	1.34E+07	-2.90E-01	-2.90E-01	2.81E+01
5	4.02E-02	4.21E+00	4.42E-01	8.50E-01	1.61E+00	-1.79E-01	5.00E+01	4.21E+00	4.13E+03	9.51E+04	-4.43E-01	-4.43E-01	6.05E+00
6	2.03E-02	5.24E+00	4.09E-01	6.62E-01	1.76E+00	2.70E-01	5.62E+01	5.24E+00	3.73E+03	7.08E+04	-2.04E-01	-2.04E-01	6.99E+00
7	1.61E-02	4.14E+00	5.10E-01	8.35E-01	1.38E+00	3.80E-02	9.43E+01	4.14E+00	8.70E+03	2.38E+05	-2.13E-01	-2.13E-01	9.03E+00
8	7.21E-	6.50E+0	4.78E-	9.52E-01	1.62E+0	-4.90	1.93E+03	6.50E+0	7.16E+0	6.83E+07	-4.51E-	-4.51	4.32E+01

	03	0	01		0	E-02		0	5		02	E-02	
9	4.55 E-02	1.12 E+02	4.08 E-02	7.66 E-01	8.73 E+00	0.00 E+00	2.02E +03	1.12 E+02	8.88 E+05	9.14E+07	0.00E +00	0.00 E+00	4.29 E+01
10	2.35 E-02	2.46 E+00	5.47 E-01	8.31 E-01	1.15 E+00	-4.08 E-02	7.69E +01	2.46 E+00	6.00 E+03	1.26E+05	4.77E-02	4.77 E-02	8.35 E+00
11	4.17 E-02	3.17 E+01	1.83 E-01	7.97 E-01	4.33 E+00	0.00 E+00	3.60E +02	3.17 E+01	7.19 E+04	3.50E+06	0.00E +00	0.00 E+00	1.71 E+01
12	4.09 E-02	1.41 E+00	6.03 E-01	9.83 E-01	8.97 E-01	8.68 E-02	7.61E +01	1.41 E+00	9.74 E+03	3.90E+05	-9.08E-02	-9.08 E-02	6.77 E+00
13	1.69 E-02	2.13 E+00	5.91 E-01	9.43 E-01	1.02 E+00	6.84 E-02	1.90E +02	2.13 E+00	2.40 E+04	8.33E+05	-1.57E-01	-1.57 E-01	1.31 E+01
14	5.96 E-03	1.83 E+00	6.34 E-01	9.95 E-01	8.98 E-01	2.07 E-02	6.12E +02	1.83 E+00	1.75 E+05	1.41E+07	1.71E-01	1.71 E-01	2.07 E+01
15	4.54 E-03	4.58 E+00	4.71 E-01	9.81 E-01	1.55 E+00	-1.40 E-02	3.66E +02	4.58 E+00	8.51 E+04	5.76E+06	6.31E-02	6.31 E-02	1.58 E+01
16	8.43 E-03	3.54 E+00	4.98 E-01	9.52 E-01	1.38 E+00	-1.76 E-02	2.47E +02	3.54 E+00	3.80 E+04	1.62E+06	1.49E-02	1.49 E-02	1.45 E+01
17	5.35 E-02	1.26 E+00	6.37 E-01	8.84 E-01	8.15 E-01	-3.56 E-01	2.31E +01	1.26 E+00	1.18 E+03	1.52E+04	-9.72E-02	-9.72 E-02	4.30 E+00
18	3.36 E-03	1.12 E+01	3.63 E-01	9.40 E-01	2.40 E+00	-3.40 E-02	4.28E +02	1.12 E+01	8.06 E+04	4.31E+06	2.47E-02	2.47 E-02	1.87 E+01
19	2.16 E-02	2.12 E+00	5.94 E-01	8.84 E-01	1.01 E+00	-2.42 E-02	1.35E +02	2.12 E+00	1.36 E+04	3.65E+05	-4.32E-01	-4.32 E-01	1.12 E+01
20	1.54 E-02	3.49 E+00	5.41 E-01	9.11 E-01	1.29 E+00	8.23 E-02	1.93E +02	3.49 E+00	2.47 E+04	8.65E+05	2.04E-01	2.04 E-01	1.32 E+01
21	2.73 E-02	6.33 E+00	5.14 E-01	9.56 E-01	1.55 E+00	-9.92 E-02	1.86E +03	6.33 E+00	6.64 E+05	6.07E+07	-1.37E-02	-1.37 E-02	4.22 E+01
22	1.38 E-02	2.29 E+00	5.71 E-01	9.41 E-01	1.08 E+00	-4.78 E-03	3.81E +02	2.29 E+00	6.38 E+04	2.79E+06	-1.60E-01	-1.60 E-01	1.90 E+01

23	2.33E-02	1.94E+00	5.88E-01	8.99E-01	9.98E-01	-2.74E-02	8.10E+01	1.94E+00	6.88E+03	1.62E+05	7.76E-02	7.76E-02	8.45E+00
24	1.33E-02	6.67E+00	4.73E-01	9.88E-01	1.70E+00	-6.86E-02	1.81E+03	6.67E+00	6.63E+05	6.18E+07	-2.42E-01	-2.42E-01	4.07E+01
25	8.05E-03	4.26E+00	4.92E-01	9.60E-01	1.46E+00	6.04E-02	1.60E+02	4.26E+00	2.65E+04	1.40E+06	9.57E-02	9.57E-02	1.03E+01

Table 7: Textural Features Extracted by Sparse GLAM for 25 images

S.No.	ASM	CONTR-AST	IDM	CORREL-ATION	DISSIMI-LARITY	ENTROPY	SUM OF SQUARES OF VARIANCE	INERTIA	CLUSTER SHADE	CLUSTER PROMINENCE	DIFFERENCE ENTROPY	SUM ENTROPY	SUM AVERAGE
1	1.11E+00	7.46E-03	7.27E-01	9.97E-01	6.32E-01	2.30E-01	8.65E+02	1.11E+00	2.62E+05	2.12E+07	-9.11E-02	-9.11E-02	2.57E+01
2	8.37E-01	9.78E-03	7.60E-01	9.98E-01	5.37E-01	-5.55E-02	1.04E+03	8.37E-01	3.31E+05	2.77E+07	-5.67E-01	-5.67E-01	2.87E+01
3	8.19E-01	8.70E-03	7.54E-01	9.98E-01	5.45E-01	1.51E-01	1.35E+03	8.19E-01	4.63E+05	4.19E+07	1.30E-01	1.30E-01	3.43E+01
4	5.97E-01	1.76E-02	7.83E-01	9.93E-01	4.60E-01	-1.67E-01	4.73E+02	5.97E-01	9.30E+04	4.84E+06	5.91E-02	5.91E-02	2.07E+01
5	7.06E-01	9.19E-02	7.48E-01	7.51E-01	5.37E-01	7.53E-02	9.11E+00	7.06E-01	2.53E+02	1.93E+03	6.63E-01	6.63E-01	2.81E+00
6	6.61E-01	8.37E-02	7.49E-01	8.64E-01	5.29E-01	-4.11E-03	1.11E+01	6.61E-01	3.91E+02	3.96E+03	9.11E-01	9.11E-01	2.97E+00
7	9.76E-01	3.65E-02	7.07E-01	9.32E-01	6.51E-01	9.69E-02	2.89E+01	9.76E-01	1.66E+03	2.63E+04	2.64E-01	2.64E-01	4.67E+00
8	3.67E+00	1.27E-02	5.92E-01	9.68E-01	1.19E+00	3.74E-02	1.66E+02	3.67E+00	2.50E+04	1.03E+06	-1.15E-01	-1.15E-01	1.04E+01

9	1.75E+00	9.29E-03	6.23E-01	9.86E-01	9.04E-01	9.28E-03	3.11E+02	1.75E+00	5.43E+04	2.50E+06	1.69E-01	1.69E-01	1.57E+01
10	4.79E-01	5.70E-02	7.89E-01	9.46E-01	4.32E-01	7.41E-02	2.18E+01	4.79E-01	1.04E+03	1.36E+04	1.68E-01	1.68E-01	4.17E+00
11	2.68E+00	7.64E-03	6.01E-01	9.83E-01	1.06E+00	3.34E-02	4.01E+02	2.68E+00	7.90E+04	4.13E+06	-1.46E-01	-1.46E-01	1.79E+01
12	7.04E-01	1.00E-01	7.25E-01	7.83E-01	5.76E-01	1.41E-01	1.37E+01	7.04E-01	4.45E+02	3.85E+03	1.31E-01	1.31E-01	3.51E+00
13	8.84E-01	3.67E-02	7.37E-01	9.71E-01	5.83E-01	-2.42E-01	8.47E+01	8.84E-01	8.10E+03	2.21E+05	-8.48E-02	-8.48E-02	8.34E+00
14	6.87E-01	9.75E-03	7.86E-01	9.98E-01	4.68E-01	2.70E-02	7.01E+02	6.87E-01	2.09E+05	1.75E+07	5.22E-02	5.22E-02	2.26E+01
15	1.39E+00	9.33E-03	6.85E-01	9.94E-01	7.46E-01	-8.14E-02	3.60E+02	1.39E+00	8.37E+04	5.67E+06	-2.20E-01	-2.20E-01	1.56E+01
16	1.24E+00	1.65E-02	6.82E-01	9.81E-01	7.31E-01	2.01E-01	1.45E+02	1.24E+00	1.87E+04	6.81E+05	-3.36E-01	-3.36E-01	1.06E+01
17	4.33E-01	1.02E-01	8.13E-01	9.30E-01	3.83E-01	2.32E-01	1.37E+01	4.33E-01	5.20E+02	5.30E+03	2.30E-01	2.30E-01	3.29E+00
18	5.28E+00	6.10E-03	5.62E-01	9.69E-01	1.40E+00	1.34E-01	4.33E+02	5.28E+00	8.90E+04	4.90E+06	4.74E-01	4.74E-01	1.87E+01
19	6.37E-01	4.75E-02	7.79E-01	9.51E-01	4.74E-01	3.79E-02	3.72E+01	6.37E-01	2.27E+03	3.81E+04	6.44E-01	6.44E-01	5.55E+00
20	1.32E+00	4.01E-02	7.20E-01	9.53E-01	6.77E-01	1.07E-01	5.17E+01	1.32E+00	4.26E+03	1.01E+05	-4.72E-02	-4.72E-02	6.14E+00
21	1.19E+00	9.14E-02	7.36E-01	9.04E-01	6.29E-01	-1.45E-01	1.52E+01	1.19E+00	9.51E+02	2.05E+04	-9.43E-01	-9.43E-01	3.00E+00
22	8.18E-01	2.64E-02	7.33E-01	9.73E-01	5.82E-01	-1.24E-02	2.26E+02	8.18E-01	2.99E+04	1.05E+06	2.06E-01	2.06E-01	1.45E+01

23	8.38E-01	4.15E-02	7.35E-01	9.42E-01	5.81E-01	-1.16E-02	2.84E+01	8.38E-01	1.67E+03	2.81E+04	1.62E-01	1.62E-01	4.61E+00
24	1.84E+00	3.04E-02	6.92E-01	9.95E-01	7.66E-01	1.35E-01	1.76E+03	1.84E+00	6.38E+05	5.84E+07	9.16E-02	9.16E-02	3.98E+01
25	1.65E+00	1.46E-02	6.93E-01	9.84E-01	7.59E-01	2.23E-02	1.68E+02	1.65E+00	2.68E+04	1.33E+06	-4.14E-01	-4.14E-01	1.07E+01

Table 8: Textural Features Extracted by Improved Sparse GLAM for 25 images

S.No.	ASM	CONTR-AST	IDM	CORREL-ATION	DISSIMI-LARITY	ENTROPY	SUM OF SQUARES OF VARIANCE	INERTIA	CLUSTER SHADE	CLUSTER PROMINENCE	DIFFERENCE ENTROPY	SUM ENTROPY	SUM AVERAGE
1	4.86E+00	7.57E-03	7.30E-01	9.97E-01	6.21E-01	2.52E-01	8.77E+02	1.07E+00	2.66E+05	2.16E+07	-2.06E-01	-2.06E-01	2.59E+01
2	6.35E+00	9.88E-03	7.61E-01	9.98E-01	5.33E-01	-5.53E-02	1.05E+03	8.26E-01	3.34E+05	2.80E+07	-5.67E-01	-5.67E-01	2.88E+01
3	9.61E+00	9.08E-03	7.60E-01	9.98E-01	5.29E-01	3.32E-01	1.39E+03	7.79E-01	4.82E+05	4.39E+07	-1.19E-02	-1.19E-02	3.49E+01
4	2.93E+00	1.80E-02	7.87E-01	9.93E-01	4.51E-01	1.30E-01	4.65E+02	5.74E-01	9.05E+04	4.66E+06	4.93E-01	4.93E-01	2.05E+01
5	6.17E-03	9.51E-02	7.52E-01	7.60E-01	5.24E-01	3.79E-02	9.12E+00	6.63E-01	2.53E+02	1.91E+03	1.69E-01	1.69E-01	2.82E+00
6	1.06E-02	8.99E-02	7.48E-01	8.54E-01	5.28E-01	-3.52E-02	1.06E+01	6.53E-01	3.59E+02	3.56E+03	9.30E-01	9.30E-01	2.92E+00
7	2.81E-02	3.65E-02	7.07E-01	9.32E-01	6.51E-01	9.69E-02	2.89E+01	9.76E-01	1.66E+03	2.63E+04	2.64E-01	2.64E-01	4.67E+00
8	1.71E-01	1.50E-02	6.20E-01	9.73E-01	1.06E+00	-1.02E-01	1.51E+02	3.03E+00	2.26E+04	9.31E+05	-1.04E-01	-1.04E-01	9.76E+00

9	6.95E-01	9.46E-03	6.28E-01	9.87E-01	8.91E-01	-5.58E-03	3.07E+02	1.72E+00	5.38E+04	2.47E+06	-6.08E-02	-6.08E-02	1.56E+01
10	1.50E-02	5.70E-02	7.89E-01	9.46E-01	4.32E-01	7.41E-02	2.18E+01	4.79E-01	1.04E+03	1.36E+04	1.68E-01	1.68E-01	4.17E+00
11	8.27E-01	7.71E-03	6.03E-01	9.83E-01	1.06E+00	6.08E-03	3.99E+02	2.75E+00	7.90E+04	4.12E+06	-1.15E-01	-1.15E-01	1.78E+01
12	7.03E-03	1.00E-01	7.25E-01	7.83E-01	5.76E-01	1.41E-01	1.37E+01	7.04E-01	4.45E+02	3.85E+03	1.31E-01	1.31E-01	3.51E+00
13	1.93E-01	3.77E-02	7.44E-01	9.72E-01	5.67E-01	-1.44E-01	8.39E+01	8.45E-01	7.98E+03	2.17E+05	8.52E-02	8.52E-02	8.30E+00
14	2.11E+01	1.06E-02	7.94E-01	9.98E-01	4.48E-01	8.79E-02	6.50E+02	6.40E-01	1.90E+05	1.57E+07	2.18E-02	2.18E-02	2.16E+01
15	1.07E+01	9.27E-03	6.88E-01	9.94E-01	7.37E-01	1.31E-01	3.70E+02	1.36E+00	8.69E+04	5.94E+06	2.67E-01	2.67E-01	1.58E+01
16	4.60E-01	1.66E-02	6.83E-01	9.81E-01	7.27E-01	2.01E-01	1.45E+02	1.23E+00	1.88E+04	6.86E+05	-3.34E-01	-3.34E-01	1.06E+01
17	6.95E-03	1.02E-01	8.13E-01	9.30E-01	3.83E-01	2.32E-01	1.37E+01	4.33E-01	5.20E+02	5.30E+03	2.30E-01	2.30E-01	3.29E+00
18	6.41E-01	6.45E-03	5.76E-01	9.70E-01	1.35E+00	1.34E-01	4.31E+02	5.09E+00	8.92E+04	4.91E+06	2.88E-01	2.88E-01	1.86E+01
19	3.79E-02	4.75E-02	7.79E-01	9.51E-01	4.74E-01	3.79E-02	3.72E+01	6.37E-01	2.27E+03	3.81E+04	6.44E-01	6.44E-01	5.55E+00
20	1.05E-01	4.01E-02	7.20E-01	9.53E-01	6.77E-01	1.07E-01	5.17E+01	1.32E+00	4.26E+03	1.01E+05	-4.72E-02	-4.72E-02	6.14E+00
21	4.81E-02	9.07E-02	7.35E-01	9.03E-01	6.32E-01	-1.44E-01	1.53E+01	1.19E+00	9.58E+02	2.06E+04	-9.42E-01	-9.42E-01	3.02E+00
22	3.70E-01	2.64E-02	7.33E-01	9.73E-01	5.82E-01	-1.24E-02	2.26E+02	8.18E-01	2.99E+04	1.05E+06	2.06E-01	2.06E-01	1.45E+01

23	4.39E-02	4.09E-02	7.34E-01	9.43E-01	5.83E-01	-7.19E-03	2.89E+01	8.43E-01	1.71E+03	2.88E+04	1.49E-01	1.49E-01	4.64E+00
24	3.45E-01	3.24E-02	6.96E-01	9.93E-01	7.70E-01	1.38E-01	1.82E+03	1.95E+00	6.62E+05	6.08E+07	2.39E-01	2.39E-01	4.09E+01
25	1.03E+01	1.46E-02	6.94E-01	9.84E-01	7.58E-01	2.25E-02	1.68E+02	1.65E+00	2.68E+04	1.33E+06	-4.14E-01	-4.14E-01	1.07E+01

Table 9: Time taken by traditional Fixed Window, Sparse and Improved Sparse ROIs for GLCM and GLAM

S. No.	GLCM (time in sec.)		GLAM (time in sec.)		
	Fixed window	Sparse	Fixed window	Sparse	Improved Sparse
1	143.5852	2.957019	260.482	1352.123	2.955628
2	145.2275	4.448406	268.2461	1882.818	3.904705
3	164.9035	6.460743	298.6097	2054.91	5.458111
4	188.369	2.515627	302.0604	1113.971	1.799991
5	191.4232	0.023025	423.9706	0.044355	0.005023
6	220.9597	0.022237	406.6209	0.225772	0.009506
7	205.5852	0.148843	314.8274	5.053507	0.024971
8	275.5019	0.586013	387.4853	34.73158	0.131029
9	259.5557	0.319164	305.5605	357.4795	0.45015
10	271.5191	0.036462	243.8996	0.369994	0.011707
11	262.0639	0.069719	315.1551	385.7937	0.534382
12	271.3971	0.04036	295.1555	0.008339	0.004593
13	262.7504	0.205182	395.8247	65.91151	0.129517
14	359.9116	12.99278	440.8307	7232.642	12.87329
15	352.7992	6.321749	269.7271	3854.157	6.409962
16	348.723	0.451405	235.8287	187.4673	0.315695
17	340.0404	0.009062	154.3511	0.037845	0.005808
18	347.0623	0.574854	157.4065	154.0435	0.437949
19	346.0856	0.07669	457.7204	14.92219	0.027928
20	342.2221	0.119756	403.0937	19.70896	0.054855
21	341.8415	0.377752	286.6165	1.420357	0.035128
22	340.6856	0.339082	258.2849	97.33222	0.250247
23	339.8158	0.071054	341.4162	5.57726	0.033975
24	339.6045	0.295753	337.6345	26.3731	0.281756
25	338.9166	5.626249	449.4328	4183.488	6.043973

The confusion matrix in all seven cases under various schemes such as Fixed Window, Sparse, Improved Sparse within that GLCM and GLAM varying percent of training data set is computed, and the elements are depicted in Table 10.

Table 10: Confusion Matrix for seven classes under various schemes

%	Method	GLCM/GLAM	1				2				3				4				5				6				7			
			TP	FP	FN	TN	TP	FP	FN	TN	TP	FP	FN	TN	TP	FP	FN	TN	TP	FP	FN	TN	TP	FP	FN	TN	TP	FP	FN	TN
90	F. W.	GLCM	2	0	1	33	3	0	0	33	2	0	0	34	2	0	0	34	3	0	0	33	2	0	0	34	21	1	0	14
		GLAM	3	0	0	33	3	0	0	33	2	0	0	34	2	0	0	34	2	0	1	33	2	0	0	34	21	1	0	14
	SPARSE	GLCM	3	0	0	33	3	0	0	33	2	0	0	34	2	0	0	34	2	0	1	33	2	0	0	34	21	1	0	14
		GLAM	2	0	0	90	2	0	0	90	2	0	0	90	5	0	0	87	8	0	0	84	3	0	0	89	70	0	0	22
	Improved Sparse	GLAM	2	0	0	23	2	0	0	23	2	0	0	23	2	0	0	23	2	0	0	23	1	0	0	24	14	0	0	11
80	F. W.	GLCM	4	0	1	1	5	0	0	61	3	1	1	61	2	0	1	63	5	0	0	61	3	0	0	63	40	3	1	22
		GLAM	3	0	2	61	3	0	2	61	4	0	0	62	3	0	0	63	5	0	0	61	2	0	1	63	41	5	0	20
	SPARSE	GLCM	4	0	1	61	4	0	1	61	3	0	1	62	2	0	1	63	5	0	0	61	3	0	0	63	41	4	0	21
		GLAM	3	0	1	43	4	0	0	43	3	0	1	43	3	0	0	44	3	0	0	44	1	0	0	46	28	2	0	17
	Improved Sparse	GLAM	3	0	1	43	4	0	0	43	4	0	0	43	3	0	0	44	2	0	1	44	1	0	0	46	28	2	0	17

70	F. W.	GL CM	5	0	3	92	4	0	3	93	5	0	1	94	5	0	0	95	5	0	2	93	4	0	1	95	62	10	0	28
		GL AM	5	0	3	92	6	0	1	93	4	0	2	94	3	0	2	95	6	0	1	93	4	0	1	95	62	10	0	28
	SPARSE	GL CM	6	0	2	92	4	0	3	93	4	0	2	94	4	0	1	95	5	0	2	93	4	0	1	95	62	11	0	27
		GL AM	3	0	3	64	6	0	0	64	5	0	1	64	3	0	1	66	3	0	1	66	1	0	1	68	42	7	0	21
	Improved Sparse	GL AM	5	0	1	64	5	0	1	64	3	0	3	64	4	0	0	66	4	0	0	66	2	0	0	68	42	5	0	23
60	F. W	GL CM	6	1	4	119	7	1	2	120	5	0	3	122	4	0	2	124	6	0	3	121	2	0	4	124	80	18	2	30
		GL AM	6	0	4	120	5	0	4	121	4	0	4	122	3	0	3	124	5	0	4	121	5	0	1	124	82	20	0	28
	SPARSE	GL CM	4	0	6	120	6	0	3	121	6	0	2	122	4	0	2	124	5	0	4	121	5	0	1	124	82	18	0	30
		GL AM	4	0	3	85	5	0	3	84	6	0	2	84	2	0	3	87	4	0	2	86	1	0	1	90	56	14	0	22
	Improved Sparse	GL AM	5	0	2	85	6	0	2	84	6	0	2	84	5	0	0	87	3	0	3	86	1	0	1	90	56	10	0	26

As previously discussed, Region of Interest (ROI) is determined using three methodologies: Fixed Window, Sparse, and Improved Sparse. For each method, feature extraction is performed using Gray-Level Co-occurrence Matrix (GLCM) and Gray-Level Aura Matrix (GLAM) where applicable. The performance of the Multi-class Support Vector Machine (SVM) is evaluated based on Accuracy, Sensitivity, and Precision. The relative performance of these methods is presented in Table 11 with 90%, 80%, 70% & 60% training data and in Table 12.

Table 11: Relative performance of multi-SVM w.r.t. various methods of Feature Extractions and Data Structures.

% of training	Method	Type for Feature Extraction	CALC	CIRC	SPIC	MISC	ARCH	ASYM	NORM	AVERAGE
90	Fixed Window	GLCM A	100	100	100	100	100	100	95.5	97.2
		P	100	100	100	100	100	100	95.4	99.34
		S	66.6	100	100	100	100	100	100	95.23
		GLAM A	100	100	100	100	100	100	95.5	97.2
		P	100	100	100	100	100	100	95.4	99.34
		S	100	100	100	100	66.6	100	100	95.22
	Sparse	GLCM A	100	100	100	100	100	100	95.5	97.2
		P	100	100	100	100	100	100	95.4	99.34
		S	100	100	100	100	66.6	100	100	95.23
		GLAM A	100	100	100	100	100	100	100	100
		P	100	100	100	100	100	100	100	100
		S	100	100	100	100	100	100	100	100
	Improved sparse	GLAM A	100	100	100	100	100	100	100	100
		P	100	100	100	100	100	100	100	100

		S								
80	**Fixed Window**	GLCM A	100	100	100	100	100	100	91.1	93.90
		P	100	100	75	100	100	100	93.00	95.43
		S	80	100	75	66.6	100	100	97.5	88.44
		GLAM A	100	100	100	100	100	100	89.10	98.44
		P	100	100	100	100	100	100	89.10	98.44
		S	60	60	100	100	100	66.6	100	83.80
	Sparse	GLCM A	100	100	100	100	100	100	91.10	93.90
		P	100	100	100	100	100	100	91.10	98.73
		S	80	80	75	66.6	100	100	100	85.94
		GLAM A	100	100	100	100	100	100	93.30	95.70
		P	100	100	100	100	100	100	93.30	99.04
		S	75	100	75	100	100	100	100	92.85
	Improved sparse	GLAM A	100	100	100	100	100	100	93.30	95.70
		P	100	100	100	100	100	100	93.30	99.04
		S	75	100	100	100	66.6	100	100	91.66

Performance measures: (A-Accuracy, P-precision and S-Sensitivity)

Table 11: (contd.): Relative performance of multi-SVM w.r.t. various methods of Feature Extractions and Data Structures.

% of Training	Method	Type for Feature Extraction	CALC	CIRC	SPIC	MISC	ARCH	ASYM	NORM	AVERAGE
70	Fixed Window	GLCM A	100	100	100	100	100	100	86.1	90
		P	100	100	100	100	100	100	86.1	98.01
		S	62.5	57.1	83.3	100	71.4	80	100	79.18
		GLAM A	100	100	100	100	100	100	86.1	90
		P	100	100	100	100	100	100	86.1	98.01
		S	62.5	85.7	66.6	60	85.7	80	100	77.21
	Sparse	GLCM A	100	100	100	100	100	100	89.9	93
		P	100	100	100	100	100	100	84.9	97.84
		S	75	57.1	66.6	80	71.4	80	100	75.73
		GLAM A	100	100	100	100	100	100	85.7	90
		P	100	100	100	100	100	100	100	100
		S	50	100	83.3	75	75	50	100	76.18
	Improved sparse	GLAM A	100	100	100	100	100	100	89.4	92.9
		P	100	100	100	100	100	100	89.3	98.47
		S	83.3	83.3	50	100	100	100	100	88.09
60	Fixed Window	GLCM A	100	100	100	100	100	100	81.6	84.6
		P	85.7	85.7	100	100	100	100	81.6	93.54
		S	60	77.7	62.5	66.6	66.6	33.3	97.5	66.31
		GLAM A	100	100	100	100	100	100	81.6	84.6
		P	100	100	100	100	100	100	80.3	97.19
		S	60	55.5	50	50	55.5	83.3	100	64.9
	Sparse	GLCM A	100	100	100	100	100	100	82.8	86.9
		P	100	100	100	100	100	100	82	97.43
		S	40	66.6	75	66.6	55.5	83.3	100	69.57

		GLAM	A	100	100	100	100	100	100	80	84.8
			P	100	100	100	100	100	100	80	97.14
			S	57.1	62.5	75	40	66.6	50	100	64.45
	Improved sparse	GLAM	A	100	100	100	100	100	100	84.8	97.82
			P	100	100	100	100	100	100	84.8	97.82
			S	71.4	75	75	100	50	50	100	74.48

Performance measures: (A-Accuracy, P-precision and S-Sensitivity)

Overall conclusions based on average performances: Here we present overall conclusions based on the last column of Table 11 which contains the average of various performance measures viz., Accuracy, Precision and Sensitivity under various schemes. These values are reproduced here for ready reference. Table 12 is created to represent average performance of Accuracy, Precision and Sensitivity under various schemes fixing GLCM.

Table 12: Maximum Average performance measures under Fixed Window and Sparse fixing GLCM

	A		**P**		**S**	
	W	**S**	**W**	**S**	**W**	**S**
90%	97.2	97.2	99.34	99.34	95.23	95.23
80%	93.90	93.90	95.43	98.73	88.44	85.94
70%	90.00	93.00	98.01	97.84	79.18	75.73
60%	84.60	86.90	93.54	97.43	66.31	69.57

From the above table it can be seen that the maximum Accuracy of 97.2% has occurred when taken 90% of training data set under Fixed window and Sparse indicating that both are equally good. Hence if one is considering 90% of training data set and performance measure Accuracy with respect to GLCM then Fixed Window or Sparse schemes are performing equally good. Similar observations can be made w.r.t Precision and Sensitivity.

Similarly, the average performances of various measures fixing GLAM are presented in Table 13.

Table 13: Maximum Average performance measures under Fixed Window, Sparse and Improved Sparse fixing GLAM

	A			**P**			**S**		
	W	**S**	**I**	**W**	**S**	**I**	**W**	**S**	**I**
90%	97.2	100	100	99.34	99.34	100	95.22	100	100
80%	98.44	95.70	95.70	98.44	99.04	99.04	83.80	92.85	91.66
70%	90.00	90.00	92.90	98.01	100	98.47	77.21	76.18	88.09

60%	84.60	84.80	97.82	97.19	97.14	97.82	64.90	64.45	74.48

From the above table it can be seen that the maximum Accuracy of 100% has been observed when taken 90% of training data set under Sparse and Improved Sparse inferring that both are equally good. Hence if one is considering 90% of training data set and performance measure Accuracy with respect to GLAM then Sparse or Improved Sparse schemes are performing equally good. Similar observations can be made w.r.t. Precision and Sensitivity.

Denoting (1), (2), (3), and (4) respectively for 90%, 80%, 70% and 60% training data set taken at random from each class. The maximum value of various performance measures are taken into Table 14 and are as follows.

Table 14: Maximum Relative performance of various measures under various schemes fixing GLCM and GLAM

	GLCM			**GLAM**			
Max	**W**	**S**	**Min**	**W**	**S**	**I**	**Min**
A	$97.2_{(1)}$	$97.2_{(1)}$	$97.2_{(W1 \text{ or } S1)}$	$98.4_{(2)}$	$100_{(1)}$	$100_{(1)}$	$98.4_{(W2)}$
P	$99.34_{(1)}$	$99.34_{(1)}$	$99.34_{(W1 \text{ or } S1)}$	$99.34_{(1)}$	$100_{(1)(3)}$	$100_{(1)}$	$99.34_{(W1)}$
S	$95.23_{(1)}$	$95.23_{(1)}$	$95.23_{(W1 \text{ or } S1)}$	$95.22_{(1)}$	$100_{(1)}$	$100_{(1)}$	$95.22_{(W1)}$

In the above table A, P and S respectively denote Accuracy, Precision and Sensitivity. W, S and I respectively represent Fixed Window, Sparse and Improved Sparse.

Notations used in Table 15 to present the final conclusions are explained below.

S-AM : Sparse GLAM
I-AM : Improved Sparse GLAM
W-AM : Fixed Window GLAM
W-CM : Fixed Window GLCM
S-CM : Sparse GLCM

Table 15: The preferences of order of choice of different data structures for the classification using multi SVM.

	Accuracy	**Precision**	**Sensitivity**
90%	S-AM/I-AM, W-AM/W-CM/S-CM	I-AM, W-CM/W-AM/S-CM/S-AM	S-AM/I-AM, W-CM/S-CM, W-AM
80%	W-AM, S-AM/I-AM, W-CM/S-CM	S-AM/I-AM, S-CM, W-AM, W-CM	S-AM, I-AM, W-CM, S-CM, W-AM
70%	S-CM, I-AM, W-CM/W-AM/S-AM	S-AM, I-AM, W-CM, W-AM, S-CM	I-AM, W-CM, W-AM, S-AM, S-CM
60%	I-AM, S-CM, S-AM, W-CM/W-AM	I-AM, S-CM, W-AM, S-AM, W-CM	I-AM, S-CM, W-AM, S-AM, W-CM

Recommendations

Based on Table 16 the following recommendations are made from the practitioners point of view.

When the performance measure of Accuracy is concern with 90% of training data set then the order of preference of feature extraction to attain the maximum possible Accuracy is GLAM with Sparse (S-AM) or Improved Sparse GLAM (I-AM), Fixed Window GLAM (W-AM) or Fixed Window GLCM (W-CM) or Sparse GLCM (S-CM).

Similar interpretation can be made for 80%, 70% and 60% training data set w.r.t Accuracy.

When the performance measure of Precision is considered with 90% training data set then the order of preference is Improved Sparse GLAM (I-AM) and the next is Fixed Window GLAM (W-AM) or Sparse GLCM (S-CM) or Sparse GLAM (S-AM).

Similar interpretation can be made for 80%, 70% and 60% training data set w.r.t. Precision.

When the performance measure of Sensitivity is considered with 90% training data set then the order of preference is Sparse GLAM (S-AM) or Improved Sparse GLAM (I-AM), Fixed Window GLCM (W-CM) or Sparse GLCM (S-CM), Fixed Window GLAM (W-AM).

Similar interpretation can be made for 80%, 70% and 60% training data set w.r.t Sensitivity.

From the above observations in majority of the cases Improved Sparse which is proposed in this thesis appears to be better choice when compared to Fixed Window or Sparse.

VII. References

1. Annavarapu Jagadish Kumar, “Machine Learning Tools for Medical Image Diagnostics”, Unpublished Ph. D thesis in Statistics, degree awarded by Acharya Nagarjuna University, Guntur, Andhra Pradesh, May 2019.
2. Annavarapu J.K., Dattatreya Rao A.V., Karteeka Pavan K., 2018, “Region of Interest (ROI) And Feature Extraction From The Images of mammograms For the diagnosis of Breast cancer”. The International Journal Research Publication’s, Research Journal of Science & IT management, ol.No.07, pp.9-28.
3. Annavarapu J.K., Dattatreya Rao A.V., Karteeka Pavan K., 2018, “Pre-Processing of Images For the Diagnosis of Breast Cancer”, The International Journal Research Publication’s, Research Journal of Science & IT Management, Vol.No.07, pp.1-8.
4. Kanchan Lata Kashyap, Koushlendra Kumar Singh, Manish Kumar Bajpai, Pritee Khanna, 2017. “Fractional Order Filter based Enhancement of Digital Mammograms”. Proceedings of the World Congress on Engineering and Computer Science. Vol.1.
5. Qinghua Huang, Yaozhong Luo, Qiangzhi Zhang, 2017. “Breast ultrasound image Segmentation: a Survey. International Journal of Computer Assisted radiology and Surgery”, 12 (3).
6. Dattatreya Rao, A.V., Annavarapu, J.K., 2016, Big Data Analytics, proceedings of Conference “Emerging Trends in Statistical Research”, Narosa Publications, New Delhi, pp.31-41.
7. Chu, J., Min, H., Liu, L., & Lu, W., 2015. “A novel computer aided breast mass detection scheme based on morphological enhancement and SLIC super pixel segmentation”. Medical Physics, 42(7), 3859–3869.
8. J K Annavarapu, 2015, “Statistical Feature Selection for Image Texture Analysis”, International Research Journal of Engineering And Technology, Vol.02, pp.546-550.
9. Zyout Imad, Czajkowska Joanna, Grzegorzek Marcin (2015) “Multi-scale Textural Feature Extraction and Particle Swarm Optimization Based Model Selection for False Positive Reduction in Mammography”, Computerized Medical Imaging and Graphics, 46 pp. 95-107.
10. Antonio García-Manso, Carlos J. García-Orellana, Horacio González-Velasco, Ramón Gallardo-

Caballero & Miguel Macías Macías. 2013. “Consistent performance measurement of a system to detect masses in mammograms based on blind feature extraction”. Bio Medical Engineering OnLine,12, 2-18.