A STUDY ON DIFFERENT FEATURE EXTRACTION TECHNIQUES FOR LESION IDENTIFICATION IN MRI BREAST IMAGES

Malu G.
University of KeralaThiruvananthapuramKerala, India 691004
malu.res11@iiitmk.ac.in

Elizabeth Sherly
Indian Institute of Information Technology and Management-Kerala
Technopark, Thiruvananthapuram, Kerala India 695581
sherly@iiitmk.ac.in

Abstract This study discusses the different feature extraction techniques for lesion identification in Dynamic Contrast Enhancement - Magnetic Resonance Imaging (DCE - MRI) of Breast. In DCE- MRI, kinetic feature extraction is a popular radiological approach used by Radiologists. However, extracting more features like entropy, homogeneity, heterogeneity, and statistical features of the region of interests would enhance the accuracy of lesion diagnosis, especially in 'not sure' (plateau) cases. This paper discuss about a survey of different feature extraction techniques such as structural, statistical, model-based and transform based methods. The paper also advocates a comparative study of feature extraction using statistical and intensity time kinetic curve methods. These two features are employed in understanding the prominence of malignancy in the lesion.

Keywords: Image Processing: Feature extraction techniques, structural methods, statistical approaches, morphological methods, Dynamic Contrast Enhanced MRI, intensity time kinetic graph

I. Introduction

Textures are complex visual patterns in repetitive nature, represents the structure of the image with its characteristics like brightness, colour, arrangement, slope, size etc. The patterns are composed of sub patterns such as perceived lightness, uniformity, density, roughness, heterogeneity, regularity, linearity, frequency, directionality, randomness, fineness, smoothness, etc. [1]. The texture analysis helps to extract meaningful information and to quantify some significant characteristics by grouping the pixels of images. In this paper, we have studied various texture analysis methods and proposed a different technique in Dynamic Contrast Enhanced Magnetic Resonance Imaging(DCE-MRI) to identify breast cancer in high sensitivity cases. The feature analysis is conducted in four phases, feature extraction, feature discrimination, feature classification and feature identification. For each phase, different methods have to be employed by considering the nature of images in order to get the most suited results. In this paper, we have focused on feature extraction that is generally classified as structural, statistical, model-based and transform based. However in DCE-MRI, the kinetic feature is a prominent factor to perform texture analysis. This work explains some of the models and approaches used for feature extraction and also proposed and implemented intensity time kinetic curve method to test the malignancy. Finally, statistical features have been computed and compared to ensure the 'not sure' cases obtained from the intensity time kinetic curve method.

II. Feature Extraction

Feature extraction is the technique of extracting various properties or features of an image, based on its characteristics for further computation. Features are represented as a function with some quantifiable measurements of the texture. These features help to obtain certain significant characteristics of the object as it quantifies the properties. The extracted features are then used for feature selection and classification. The main types of features used for extraction are general features and domain specific features. General features are application independent like colour, shape, where as domain specific features are application dependent, which are more conceptual like human faces, fingerprint. The two categories of features are low-level and high level features. In low level features, original image is used for extracting the features, however, in high level features, low level features are used for further extraction[14]. Several studies have been done for the feature extraction of masses in DCE-MRI of the breast. Masses visible in DCE-MRI have a clear cut boundary while non-masses have not. Shape, margins, enhancement are some of the typical features used for the detection of masses. The malignant lesions in the mass portions can be identified with some peculiarities such as spiculation, rim

enhancement, and washout kinetic patterns. While benign mass can be identified with smooth margin, low and homogeneous enhancement and persistent kinetic patterns. The masses and its malignancy can be identified with the features like texture features, morphological characters such as shape, margin and enhancement characteristics along with the Intensity time kinetic curve[2][10]. However, using Intensity time kinetic curve method, there are the situation of the 'not sure' cases(can be malignant or benign), in which doctors are in a dilemma to arrive at proper diagnosis. In such cases, further analysis based on the texture is required. Therefore, texture feature analysis plays a significant role to ascertain the diagnosis. Selecting better and suitable feature extraction techniques are crucial for such complex systems, especially to detect masses in DCE-MRI[11][12][14][17]. In this paper, after studying different approaches to texture analysis, we have proposed intensity time kinetic curve method for identifying the suspected regions and categorized as malignant, benign and 'not sure' cases. The 'not sure' case is again identified as benign or malignant using statistical features.

2.1.Structural Methods

Structural methods for feature extraction is more suited when the texture is a well defined primitive having a symbolic description on the image. The structure represents a texture according to its local properties and organization spatially. Mathematical morphology is an efficient technique for the structural analysis. The shape of the masses in BreastMRI can be round, oval, lobular or irregular[15][16]. Therefore, in the case of breast MRI images, the shape feature may provide a better classification, which can be confirmed with statistical analysis. Some morphological techniques using for feature extraction in DCEMRI are describing here[12].

2.1.1. Normalized Radial Length and other Morphological Technique.

The normalized radial length (NRL) based feature extraction individualize lesion contours as well as the shape of the lesion. Since the shapes and contours play, a significant role in lesion detection NRL method is crucial and is defined as the sum of the Euclidean distances from the central point of mass to the points in lesion circumscribe.

The Feature, circularity of a mass is defined as

$$Circularity = \frac{Perimeter^2}{Area} \tag{1}$$

where,

Perimeter-perimeter of the mass is calculated by measuring the sum of the pixel count on the outskirt of the mass.

Area-area of the mass is calculated by counting the number of pixels inside the boundary of the lesion.

The centroid of the tumor, the Euclidean length from the centroid to the contour is required for computing the normalized radial length.

The radial length can be computed using the formula given in equation (2) .

$$Rl(i) = \sqrt{(x(i) - X_0)^2 + (y(i) - Y_0)^2} \quad i = 1, 2, 3, \ldots N \tag{2}$$

where,

X_0, Y_0-coordinate points of the centroid of the mass

x(i), y(i)-coordinate of the pixel at the ith location on the boundary

N-pixel count in the boundary

If max(Nrl(i)) is the largest the radial length value of the extracted region then the contour feature computed via the radial length measure are describing below[3].

The normalized radial length *Nrl(i)* is

$$Nrl(i) = \frac{\sqrt{(x(i) - X_0)^2 + (y(i) - Y_0)^2}}{Max(Rl(i))} \tag{3}$$

where *max*(Rl(i)) is the maximum value of radial length *Rl(i)*

2.1.1.1) Normalized Radial Length Mean(NrlMean)

NrlMean is represented by the following:

$$NrlMean = \frac{1}{N}\sum_{i=1}^{N} Nrl(i) \tag{4}$$

2.1.1.2) Normalized Radial Length Standard Deviation

The Standard deviation of the *NRL* is figure out by

$$NrlSd = \frac{1}{N-1}\sum_{i=1}^{N}(Nrl(i) - NrlMean)^2 \quad (5)$$

NrlMean and NrlSd can pinpoint very minute rim differences in the mass[3].

2.1.1.3) Normalized Radial Length Entropy(NrlEntropy)

NrlEntropy is a histogram based probabilistic method, and the formula is

$$NrlEntropy = -\sum_{k=1}^{100} P_k(\log(P_k)) \quad (6)$$

Where, $P_k = \frac{r_k}{\sum r_k}$ is the probability density of a given *rk*. The amplitude range is set as 1 to 100 to normalize the histogram. The entropy value describes both the roughness and roundness of the masses.

2.1.1.4) Normalized Radial Length Area Ratio (NrlAreaRatio)

NrlAreaRatio is a measure of percentage of mass area outside a oblique area and is defined by the x-y line plot. The NrlAreaRatio describes the perceptible contour of the mass. It is defined as

$$NrlAreaRatio = \frac{1}{Nrl\_Mean.N}\sum_{i=1}^{N}(Nrl(i) - NrlMean)$$

$$\text{Where, } Nrl(i) - NrlMean = 0 \quad \forall \quad Nrl(i) \leq NrlMean \quad (7)$$

To identify the mass boundary, a zero-crossing count is used in addition with the area ratio measure. The zero crossing count is a count of the number of times the line plot crossed the moderate radial length.

2.1.1.5) Normalized Radial Length Roughness

It is the changes or transition located at the object edges. The mass boundary is divided into some equal length parts and the roughness is calculated for all these parts. A roughness index is calculated for each cut using the following

$$NrlRoughness = \frac{1}{N}\sum_{i=1}^{N}|Nrl(i) - Nrl(i+1)| \quad (8)$$

2.1.1.6) Normalized Radial Length Sphericity

Sphericity of a lesion is calculated by

$$Nrl Sphericity = \frac{NrlMean}{NrlSd} \quad (9)$$

Other Morphological features

2.1.1.7) Perimeter length

It is the distance or length of the boundary of the mass. To calculate the perimeter, take all the adjacent pixels and find the distance between each adjacent pair of pixels around the boundary region and sum all these values.

$$Perimeter = \sum_{i=1}^{n}\sqrt{(x_{i+1} - x_i)^2 (y_{i+1} - y_i)^2} \quad (10)$$

If a mass is malignant, then it is irregular in shape and have a larger perimeter.

2.1.1.8) Area

Area feature identifies the area of a breast lesion. Different techniques and procedures can be used to find the area of an irregular shape such like mass. Pixel counting, Cavailer formula are some of the traditional methods[6]. Malignant masses frequently have a large area compared with benign tumors.

2.1.1.9) Compactness

It calculates the contour complexity against the enclosed area and is a built-in property of an object(here it is mass). The shape of an object and shape analysis can be done using this function. A mass with smooth contour have less compactness than a mass with rough contour.

$$Compactness = \frac{Perimeter^2}{4\pi Area} \quad (11)$$

2.1.1.10)Spiculation

$$Spiculation = \frac{1}{N}\sum_{i=1}^{N}|r_i - r_{i+1}|, r_{N+1} = r_1 \quad (12)$$

where, *N* - the number of pixels on the lesion contour

ri -the length of the Euclidean distance from the object's center to each of boundary pixels. Mass with spiculated margin is a suggestive of malignancy, and the high spiculation value indicates malignant lesion.

2.1.1.11) Extent

$$Extent = \frac{Area}{AreaBox} \quad (13)$$

where, *AreaBox* - area of the smallest rectangle containing the given lesion contour.

2.1.1.12) Elongation

$$Elongation = \frac{\min(V,H)}{\max(V,H)} \quad (14)$$

Where *V* and *H* are the vertical and horizontal length of the smallest rectangle containing the given lesion contour.

2.1.1.13) Solidity

$$Solidity = \frac{Area}{Area\_Convex} \quad (15)$$

Where AreaConvex - the area of the smallest convex polygon that can contain the given lesion contour. Shape values can be used to distinguish between benign and malignant tumors. If the shape is smooth, the mass is benign otherwise malignant. When solidity is close to 0, the tumor is malignant.

2.1.1.14) Normalized Radial length Circularity

$$Circularity = \frac{1}{N}\sum_{i=1}^{N} r_i - \mu_r \quad (16)$$

Where μr is the average of *ri*. If the circularity is 1 it indicates a circle.

2.1.1.15)Roundness

$$Roundness = \frac{4\pi\, Objectarea}{Pertmeter^2} \quad 17)$$

It is similar to circularity.

2.1.1.16)Eccentricity

Eccentricity is a scalar that specifies the eccentricity of the ellipse that has the same second moments as the lesion region. It is the ratio of the distance between the foci of the ellipse and its major axis length[8].

$$Eccentricity = \sqrt{1 - \frac{b^2}{a^2}} \quad (18)$$

Where *2a* - the major axis

2b-the minor axis of the ellipse

Higher the eccentricity greater the elongation of the object. if it is a decreased value then object approaches to circular shape.

2.1.1.17) Sphericity

It is used to check the roundness of an object, and this feature is an example of closeness measure of a shape. The sphericity is calculated for a particle as the ratio of the surface area of a sphere to the surface area of the particle:

$$Sphericity = \frac{\pi^{\frac{1}{3}}(6V_p)^{\frac{2}{3}}}{A_p} \quad (19)$$

Where, Vp - volume of the particle

Ap- the surface area of the particle.

If the sphericity value is 1 then it is a sphere, any particle which is not a sphere will have sphericity less than 1.

2.1.1.18)Volume

$$Volume = Area \times Thickness \quad (20)$$

For DCE-MRI the thickness is same for all images slices. So the volume can be easily calculated when the area is given.

2.2. Statistical Approaches

Statistical approaches deals with spatial distribution of local features in images, which possess consistent properties in the distribution and relationships. This allows to compute various statistical values, which can be used to describe local features and textures.

2.2.1. Gray level histogram based features

In histogram-based approach, the analysis of texture is done on the basis of the intensity values on the entire image or a part of an image. These intensity values provide statistical information. The statistical methods can be used to check the gray values. Features derived from gray level histogram based approach include different moments and are described below. Statistical methods can be classified based on

- Number of pixels – It can be further classified as first-order that deals with only one pixel. This estimates properties of values of each pixel. There can be two pixels and higher order pixels, which estimates properties of two or more pixel values[8]. There are statistical techniques like mean, standard deviation, variance, skewness etc. Co occurrence matrix is a successful method since it provides pixel values at different locations that are relative to each other.
- Moment- First moment (mean), second moment, third central moment and fourth central moment based Classification are describing here:

(First Moment based)

2.2.1.1) Mean Intensity

$$HistMean = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N} I_1(x,y)}{M \times N} \tag{21}$$

2.2.1.2) Standard Deviation

$$HistSd = \sqrt{\frac{\sum_{x=1}^{M}\sum_{y=1}^{N}(I_1(x,y) - HistMean)^2}{M \times N}} \tag{22}$$

(Second Moment based)

2.2.1.3) Variance

$$HistVariance = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}(I_1(x,y) - HistMean)^2}{M \times N} \tag{23}$$

Variance describe the similarities of the intensities in a region.

(Third Central Moment)

2.2.1.4) Skew

$$HistSkewness = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}(I_1(x,y) - HistMean)^3}{M \times N \times HistSd^3} \tag{24}$$

skew, checks whether the intensity distribution and the mean are symmetric or not. (Fourth central Moment)

2.2.1.5 Kurtosis

$$HistKurtosis = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}(I_1(x,y) - HistMean)^4}{M \times N \times HistSd^4} \tag{25}$$

kurtosis, describes the flatness of the distribution.

2.2.1.6) Autocorrelation

An autocorrelation measures the linear spatial relationship between spatial sizes of texture primitives. This approach to texture analysis is based on the intensity value concentrations on all or part of an image represented as a feature vector. Calculation of the autocorrelation matrix involves individual pixels. The set of autocorrelation coefficients are defined by the following function.

$$Autocorrelation(p,q) = \frac{MN\sum_{i=1}^{M-p}\sum_{j=1}^{N-q} f(i,j)f(i+p,j+q)}{(M-p)\times(N-q)\sum_{i=1}^{M}\sum_{j=1}^{N} f^2(i,j)} \tag{26}$$

where p and q are the positional difference in the position in the ith and jth direction, of an M X N image.

2.2.2. Co occurrence matrix based features

Sufficient number and combinations of moments are given, it is possible to reconstruct the image in some extent, but not as a whole. This is one of the disadvantages of histogram-based measurements and this method does not store any relative information about the spatial position of pixels. In 1973 Harlick developed a spatial relationship matrix which shows how often different combinations of gray level values occur in an image. The Gray Level Co-occurrence Matrix (GLCM) involves statistical sampling of the gray levels along with the distribution of intensities, position of pixels and spatial relations. For different spatial relations, different co-occurrence matrices can be produced. Before computing the feature values, the GLCM matrix should be converted to a symmetric normalized matrix. The important and popular features that are derived from GLCM are contrast, homogeneity, dissimilarity, energy and entropy and several other features that can be derived from the above features.

To compute the GLCM, the displacement as well as the angles between neighboring pixels are taken. The displacement can be horizontal, vertical and two diagonal displacement and the different orientations are 0^0, 45^0, 90^0 and 135^0.

The notations used for the equations described below.

P(i,j) -i X j th entry in a symmetric normalized gray tone spatial dependence matrix, =P(i,j)/R.

Px(i),Py(j) - ith and jth entry in the marginal probability matrix obtained by summing the rows and column of P(i,j) respectively

$$Px(i) = \sum_{j=1}^{N_g} P(i,j)$$

$$P_y(j) = \sum_{i=1}^{N_g} P(i,j)$$

N_g - Number of distinct gray levels in the quantized image.

$$\sum_{i} = \sum_{i=1}^{N_g}$$

$$\sum_{j} = \sum_{j=1}^{N_g}$$

$$P_{x+y}(k) = \sum_{i=1}^{N_g} \sum_{\substack{j=1 \\ i+j=k}}^{N_g} p(i,j), \qquad k = 2,3,\ldots\ldots,2N_g$$

$$P_{x-y}(k) = \sum_{i=1}^{N_g} \sum_{\substack{j=1 \\ |i-j|=k}}^{N_g} p(i,j), \qquad k = 0,1,\ldots\ldots,N_g - 1$$

2.2.2.1) Uniformity / Energy / Angular Second Moment (ASM) :

$$AngularSecondMoment = \sum_{i,j} \{p(i,j)\}^2 \tag{27}$$

The local homogeneity or disorders of texture in an image is identified with the help of the statistical function Angular Second Moment (ASM). When the gray level distribution has a constant or periodic intensities then the energy value will be higher. Basically this feature will tell us how uniform the texture is. The higher the Energy value, the bigger the homogeneity of the texture. The range of Energy is [0,1], where Energy is 1 for a constant image.

2.2.2.2) Contrast / Inertia of Moment (IM)

From an adjacent set of pixels contrast is defined as the difference between the highest and the lowest values. So it gives the value of local differences present in the image. For a low contrast image the GLCM concentrations are around the principle diagonal and it have a low spatial frequency.

$$Contrast = \sum_{i,j} |i-j|^2 p(i,j) \tag{28}$$

The contrast in an image is low if the neighboring pixels are very similar in their gray level. The grey level variations represent the variation of texture in an image. If the contrast value is high it represents heavy textures and for soft texture its value is low. ie the heterogeneity of an images can be measured by its contrast value. The Contrast ranges from 0 to the square of the size of the GLCM matrix. When the contrast value is zero it will be a constant image[8].

2.2.2.3) Homogeneity / Inverse Difference Moment(IDM)

Inverse Difference Moment evaluates the homogeneity of an image. The homogeneity of an image is higher if a pair of elements have a smaller gray tone difference and the GLCM concentrates along the diagonal. Which means that there are a lot of pixels with the same or very similar grey level value. It attain maximum value one when all elements in the image are the same. The range of homogeneity is [0,1]. The homogeneity and contrast are inversely proportional.

$$Homogeneity = \sum_{i,j} \frac{1}{1+(i-j)^2} p(i,j) \tag{29}$$

High homogeneity refers to textures that contain ideal repetitive structures, while low homogeneity refers to big variation in both, texture elements and their spatial arrangements. An "inhomogeneous texture" refers to an image that has almost no repetition of texture elements and spatial similarity in it is absent[8].

2.2.2.4) Dissimilarity/ Absolute Value of Differences(AVD) 1st order statistics

Variation of gray level pairs in an image is identified by this function. It is similar to contrast but the difference is the weight. The dissimilarity grows quadratically.

$$Dissimilarity = \sum_{i,j} |i-j| p(i,j) \tag{30}$$

It is expected that these two measures behave in the same way for the same texture because they calculate the same parameter with different weights. Contrast will always give slightly higher values than Dissimilarity. Dissimilarity ranges from [0,1] and obtain maximum when the grey level of the reference and neighbor pixel is at the extremes of the possible grey levels in the texture sample.

2.2.2.5) Entropy

Entropy in any system represents randomness or disorder, where in the case of texture analysis it is a measure of its spatial disorder.

$$Entropy = -\sum_{i,j} p(i,j) \log\left(p(i,j)\right) \tag{31}$$

A completely random distribution would have very high entropy because it represents chaos. Solid would have an entropy value of 0. Entropy is inversely correlated to energy.

All the features discussed are connected in certain manner. Contrast and Dissimilarity both calculate the variation of grey level pairs, but with a different weight. Homogeneity weights values by the inverse of Contrast weight, which means lower the homogeneity, higher the Contrast. Energy is actually local Homogeneity and Entropy is the opposite of Energy[8].

2.2.2.6) GLCM Mean

$$GLCMMean_i = \sum_{i,j=0}^{N-1} i\left(P_{i,j}\right) . \quad GLCMMean_j = \sum_{i,j=0}^{N-1} j\left(P_{i,j}\right) \tag{32 A}$$

GLCMMeani calculates the mean based on the reference pixels i and GLCMMeanj calculates the mean based on the reference pixels j.

2.2.2.7) Variance / Sum of squares

$$GLCMVariance_i = \sum_{i,j=0}^{N-1} P_{i,j}(i - GLCMMean_i)^2 \tag{32}$$

$$GLCMVariance_j = \sum_{i,j=0}^{N-1} P_{i,j}(j - GLCMMean_j)^2 \tag{33}$$

In *GLCM* also the variance is a measure of the dispersion of the values around the GLCMMean.

2.2.2.8) Standard Deviation

$$GLCMStd_i = \sqrt{GLCMVariance_i^2} \tag{34}$$

$$GLCMStd_i = \sqrt{GLCMVariance_i^2} \tag{35}$$

2.2.2.9) Correlation

Gray tone linear dependencies in an image is measured by Correlation. Higher correlation values indicates similar gray level regions.

$$\sum_{i,j=0}^{N-1} P_{ij} \left[\frac{(i - GLCMMean_i)(j - GLCMMean_j)}{\sqrt{(GLCMVariance_i^2)(GLCMVariance_j^2)}} \right] \tag{36}$$

where GLCMMeani,GLCMMeanj are the means and GLCMVariancei, GLCMVariancei are the variances of P(i,j).

2.2.2.10) Sum Average

$$SumAverage = \sum_{i=2}^{2N} i p_{x+y}(i) \tag{37}$$

2.2.2.11) Sum Variance

$$SumVariance = \sum_{i=2}^{2N} (i - SumAverage)^2 P_{x+y}(i) \tag{38}$$

2.2.2.12) Sum Entropy

$$SumEntropy = -\sum_{i=2}^{2N_g} P_{x+y}(i)\log\{P_{x+y}(i)\} \tag{39}$$

2.2.2.13) Difference Entropy

$$DifferenceEntropy = -\sum_{i=0}^{N_g-1} P_{x-y}(i)\log\{p_{x-y}(i)\} \tag{40}$$

2.2.2.14) Information Measures of Correlation1

$$Information\_Measures\_of\_Correlation1 = \frac{HXY - HXY1}{\max[HX, HY]} \tag{41}$$

2.2.2.15) Information Measures of Correlation2

$$Information\_Measures\_of\_Correlation2 = 1 - \exp[-2.0(HXY2 - HXY)]^{\frac{1}{2}}$$
$$HXY = -\sum_{i,j} p(i,j)\log_g(p(i,j)) \tag{42}$$

where HX and HY are entropies of px and py, and

$$HXY1 = -\sum_{i,j} p(i,j)\log_g\{p_x(i)p_y(j)\}$$
$$HXY2 = -\sum_{i,j} p_x(i)p_y(j)\log\{p_x(i)p_y(j)\}$$

2.2.2.16) Maximal Correlation Coefficient

where

$$Maximal_{Correlation_{Coefficient}} = Second_{largest_{eigenvalue of}} Q^{1/2} =$$
$$Q(i,j) = \sum_k \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)} \tag{43}$$

2.2.2.17) Cluster Shade

It is a measure of the skewness of the matrix and is believed to gauge the perceptual concepts of uniformity. If cluster shade is high, the image is asymmetric.

$$ClusterShade = \sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} (i + j - U_x - U_y)^3 P(i,j) \tag{44}$$

2.2.2.18) Cluster Prominence

It also measures asymmetry. When the cluster prominence value is high, the image is not much symmetric.

If cluster prominence value is less, then the mean values in the GLCM matrix is high.

$$ClusterProminence = \sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} (i + j - U_x - U_y)^4 P(i,j) \tag{45}$$

Small value in cluster prominence represents small variation in gray scale values[9].

2.2.3.Run-length Matrix

On the basis of the runlength of Image's gray level textures are characterized. Galloway , Chu et. al. and Dasarathy and Holder introduced different runlength [3][4] matrices as feature representatives. For a given image, suppose the run-length matrix p(i,j) is defined as the number of runs with pixels of gray level i and run-length j. Let M be the number of gray levels and N be the maximum run length. The run-length matrix is a way of searching the image, always across a given direction, for runs of pixels having same grey- level values. The following three matrices define the traditional runlength features[7].

2.2.3.1) Gray Level Run-Length Pixel Number Matrix:

$$R_p(i,j) = p(i,j).j \tag{46}$$

Each element of the matrix represents the number of pixels of run-length j and gray level i.

2.2.3.2) Gray-Level Run-Number Vector

$$R_g(i) = \sum_{j=1}^{N} p(i,j) \tag{47}$$

This vector represents the sum distribution of the number of runs with gray level i.

2.2.3.3) Run-Length Run-Number Vector

$$R_r(j) = \sum_{i=1}^{M} p(i,j) \tag{48}$$

This vector represents the sum distribution of the number of runs with run length j. Galloway introduced five original features of run-length statistics and defined as follows.

i) Short Run Emphasis (SRE)

$$SRE = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{p(i,j)}{j^2} = \frac{1}{n_r}\sum_{j=1}^{N}\frac{p_r(j)}{j^2} \quad (49)$$

ii) Long run emphasis

$$LRE = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{p(i,j).j^2}{j^2} = \frac{1}{n_r}\sum_{j=1}^{N} p_r(j)\, j^2 \quad (50)$$

iii) Gray Level Nonuniformity

$$GLN = \frac{1}{n_r}\sum_{i=1}^{M}\left(\sum_{j=1}^{N} p(i,j)^2\right) = \frac{1}{n_r}\sum_{j=1}^{M} p_g\, i^2 \quad (51)$$

iv) RunPercentage(RP)

$$RP = \frac{n_r}{n_p} \quad (52)$$

where nr is the total number of runs and np is the number of pixels in the image

v) Low Gray Level Run Emphasis

$$GRE = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{p(i,j)}{i^2} = \frac{1}{n_r}\sum_{j=1}^{M}\frac{p_{g(i)}}{i^2} \quad (53)$$

vi) High Gray Level Run Emphasis

$$HGRE = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N} p(i,j)i^2 = \frac{1}{n_r}\sum_{j=1}^{M} p_r(j)i^2 \quad (54)$$

vii)Short Run Low Gray Level Emphasis (SRLGE)

$$SRLGE = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{p(i,j)}{i^2 j^2} \quad (55)$$

viii) Short Run Length Gray Level Emphasis(SRLGE)

$$SRLGE = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{p(i,j).i^2}{j^2} \quad (56)$$

ix) Long Run Length Gray-Level Emphasis(LRLGE)

$$LRLGE = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{p(i,j).j^2}{i^2} \quad (57)$$

x) Long Run High GrayLevel Emphasis (LRHGE)

$$LRHGE = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N} p(i,j).i^2\, j^2 \quad (58)$$

2.3. Model Based Methods

Here texture in an image is represented using sophisticated mathematical models such as fractal or stochastic and the model parameters are estimated and used for the analysis. But the drawback of these models are that it have computational complexity during the analysis.

Images are subdivided into blocks and for each block fractal dimension is calculated. Fractal dimension D measures the roughness of the block. Since fractal dimension of a tumor involved region falls within a certain region, D could be used to locate the possible cancerous region. The area of fractal surface is

$$FractalArea_r = Kr^{2-D} \quad (59)$$

where r is the ruled area

K is the scaling constant

D fractal dimension indication of the roughness of a given region.

Parameter D is computed by the Blanket Method.

2.4.The transform methods

The texture properties of the image may be analyzed in different space, such as the frequency or the scale space. These methods are based on the Fourier, Gabor or Wavelet transform, among wavelet transform is the popular one for Breast MRI.

2.4.1.Wavelets

Wavelet transform is a method to convert the signal from time domain to frequency domain, which can reveal the characterizing features of the enhancement kinetics of tissues.

Wavelets represent a technique that analyzes the frequency content of that image. This analysis yields a set of wavelet coefficients corresponding to different scale and to different frequency directions. When computing the wavelet transform of an image, we associate each pixel as a set of numbers(wavelet coefficients) which characterize the frequency content of the image at that point over a set of scale. From these coefficient we compute a set of scale[10].

2.5 Kinetic Features

In addition to the above feature extraction, Dynamic Contrast Enhanced MRI has an important feature, called Kinetic features. Nowadays, radiologist has been used this method for diagnosis which is semi-automated. The heterogeneity of lesions in Breast MRI can be easily identified with the help of Intensity Time Kinetic Graph method. For a pre-defined region of interest graph plots the enhancement of pixel intensities over different time intervals. The visualization of curve pattern provides the degree of enhancement of the lesion.

Let S(t)= {S(0),S(1),S(2),…..,S(5)} are the signal intensities at different time intervals T0,T1, ……Tn

T_0 is the time before contrast injection, $T_1,T_2,T_3 \ldots T_n$ are the different time intervals while the contrast agent spreads through the body.

2.5.1)Initial Enhancement

The initial signal increase from the pre-contrast measurement to the maximum value in time T_3 is

$$Initial\ Enhancement = \frac{\max_{t=1:3} S(t) - S(0)}{S(0)} \tag{60}$$

where S(0) is the signal intensity at the pre-contrast frame(y=0)

2.5.2) Post initial enhancement

Post initial behavior of the signal curve from the maximum peak in T_3 time to T_n time.

$$Post - Initial\ Enhancement(PIE) = \frac{S(5) - \max_{t=1:3} S(t)}{\max_{t=1:3} S(t)} \tag{61}$$

2.5.3)Signal Enhancement Ratio

Characterise the post-initial behaviour of the signal curve which is a measure of washout, incorporates both signal change in the initial and the post initial phase relative to the pre-contrast signal measurement.

$$Signal\ Enhancement\ ratio(SER) = \frac{\max_{t=1:3} S(t) - S(0)}{S(5) - S(0)} \tag{62}$$

2.5.4)Minimum enhancement parameter

The minimum enhancement of the curve is calculated using the formula

$$MinEnhancement = \frac{b - SI_{min}}{b} \tag{63}$$

where, b is the baseline signal intensity value

SI_{min} is the signal intensity value at the minimum point during the first three seconds after contrast injection of the agent.

2.5.5)Wash In Slope

$$WashInSlope = \frac{SI_{min} - b}{b(T_{min} - T_0)} \tag{64}$$

where T_{min} is the time of the occurrence of SI_{min}

T_0 is time to onset of the contrast agent.

Time to onset was found automatically by calculating the standard deviation of the baseline points and the points surrounding potential 'time to onset' points. When the standard deviation of the neighborhood of a given time point was more than three times larger the standard deviation of the baseline, that time point was denoted ‘time to onset'.

2.5.6)Washout slope

$$WashOutSlope = \frac{b_1 - SI_{min}}{b_1(T_1 - T_{min})} \quad (65)$$

where b1 - the average of intensities at frames 30 to 40

T_1 was time to return to base line, calculated similarly as T_0.

2.5.7) Change in base line intensity

Change in baseline intensity is calculated as follows:

$$Change\ in\ base\ line\ intensity = \frac{b - b1}{b} \quad (66)$$

2.5.8) Time to peak

$$TimeToPeak = T_{min} - T_0 \quad (67)$$

The features like auto correlation, contrast, cluster prominence, cluster shade, correlation, homogeneity, sum of squares shows complexity $O(n^2)$, sum average sum variance, sum entropy, solidity, dissimilarity, difference entropy etc. shows complexity O(n).

III. Lesion Detection using Kinetic Curve and Statistical Features

The suspected regions can be easily categorized as benign or malignant using persistent and washout characteristics in the Intensity Time Kinetic Graph method. In order to ascertain the categorization of lesion by using Time Kinetic Graph method, we have tried out various statistical feature extraction and also used one of the structural extraction technique called normalized radial length feature. Though the result obtained from the structural and statistical features are promising for certain features, but depending on the nature of images and its properties, the range of statistical feature values varies. Therefore, we have formulated a new technique by combining the statistical and Intensity Time Kinetic Graph method. This helped us to find out the range of feature values for malignant and benign lesions for a particular patient.

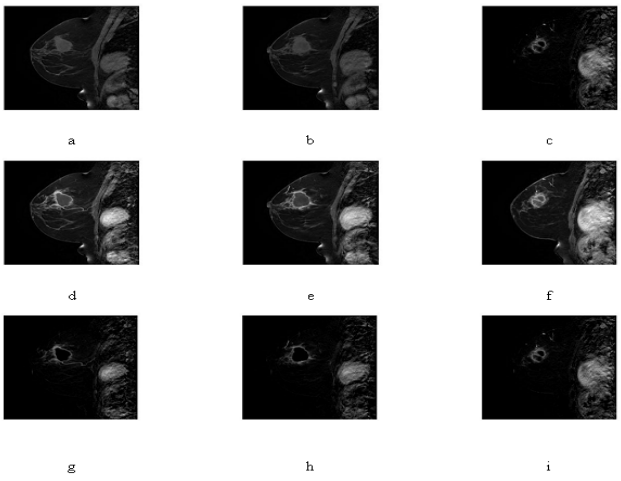

Fig. 1. a, b, c- images before contrast injection, d, e, f- images after contrast injection, g, h, i- images after, subtracting the base images with the contrast injected images"[5].

Here images consists of a particular lesion type having irregular shape, spiculated margin and rim enhancement is taken for the study (fig.1). The suspected areas are segmented into three by identifying the regions and the neighboring regions which comes under the persistent, plateau and washout curves of kinetic graph. These three sub segments consists of malignant area which comes under the washout curve pattern, benign area which comes under persistent curve pattern and the area which is not able to categorize as malignant or benign and it is named as 'not sure' and it comes under the plateau curve pattern of the Intensity time kinetic graph(Table 1). The statistics and normalized radial length features are computed for the entire suspected region, and the three segmented regions separately (Table 2.). All these sub segments show different feature values which show the peculiarities of benign region, malignant region, plateau region. Fig. 2 shows the distribution of the feature values.

Table1. Extracted areas – Entire suspected area, benign, malignant and 'not sure' areas.

Patient Name	Mass	Benign Area	Malignant Area	Not sure Case
Patient 1				
Patient 2				
Patient 3				
Patient 4				

Normally, the feature value range is standardized in a vague manner by saying the value near to zero or near to one or value is in-between two numeric values. This is because different images have different level of intensity values and type of textures, it is difficult to set a numeric range. By sub segmenting and then finding the feature values of each malignant, benign and 'not sure' case areas helps to identify the range of feature values of a particular malignant image. Then from previous studies we are tried to generalize feature values, such as the feature value is greater for malignant area than benign area or vice versa.

Here images are taken from Regional Cancer Centre, Thiruvananthapuram. 187 images which consists of suspected areas are used for the analysis. Among the images, 52 images are of a particular type and have irregular shape, spiculated margin and rim enhancement which is taken for the study. fig 1 shows some sample images.

Table 2 shows some feature values of the lesion type irregular shape, spiculated margin and rim enhancement of different patients P1, P2,P3 and P4.

	Patient	Corre1	Dissi	Energy	Homo	Corre2
Benign	P 1	0.5271	1.49	0.3874	0.8137	0.5271
	P 2	0.4618	1.63	0.386	0.7953	0.4618
	P 3	0.4219	1.57	0.4677	0.8032	0.4219
	P 4	0.5714	1.49	0.3347	0.8134	0.5714
Malignant	P 1	0.4180	1.64	0.4161	0.7941	0.4180
	P 2	0.3288	1.76	0.4355	0.7792	0.3288
	P 3	0.1615	2.88	0.2794	0.6388	0.1615
	P 4	0.5070	1.69	0.3253	0.7881	0.5070
Not Sure	P 1	0.2989	1.77	0.4485	0.7777	0.2989
	P 2	0.2723	1.79	0.4580	0.7761	0.2723
	P 3	0.1615	2.88	0.2794	0.6388	0.1615
	P 4	0.4158	1.92	0.3309	0.7597	0.4158

Observations

In the study the features such as auto correlation, cluster prominence, correlation1, correlation 2, homogeneity, sum of squares, circularity, sum average, sum variance, sum entropy, normalized radial length mean, solidity, circularity, roundness, histogram mean, histogram standard deviation show high feature values for benign area than malignant area and the plateau area shows comparatively less values. The contrast, cluster shade, dissimilarity, difference variance, difference energy, energy, max probability, Nrl entropy, roughness, histogram skewness and perimeter show high feature values for malignant portions than benign areas. Here the feature values under the plateau region area is larger than the other two.

In our results the malignant portions shows high contrast values when compared to benign regions which is an indication of malignancy. But, the area under the plateau region shows highest contrast values because contrast is the difference between the intensity values. The plateau region consists of malignant as well as benign regions their intensity differences is high. Uniformity shows the gray level distribution has a constant or periodic intensities. Here malignant images shows less uniformity and its value are near to 0. High dissimilarity indicate malignancy of the tissue. High entropy value indicates malignancy since it represents complex texture. Statistical entropy and normalized radial length entropy is calculated in our study. Low solidity value which approaches to zero is an indication malignancy. Lower energy value represents benign image. Higher correlation

values represents the benign regions. Cluster shade represents the asymmetry of the image and its value is high for malignant lesion. Small value in cluster prominence represents small variation in gray scale values and its indicates benign region. All these features describes either malignant region or benign region is greater, so the correctness can be checked.

The features describing below are shape based features and have no relevance of segmenting it as benign, malignant or 'not sure' and the entire image is used for the analysis. The feature NrlAreaRatio shows the perceptible contour of the masses. For benign masses the contour is smooth and is prominent. So for the malignant portions this feature values shows less value than the benign portions. Spiculation, circularity, roundness, elongation, extent, sphercity etc. describe the lesion shape. High spiculation indicated the malignancy of a lesion. Masses are round in shape, so the circularity, roundness, eccentricity, sphericity, etc. are used to detect this. But smooth round shape indicates the benign mass and its value approached to 1. Sphercity value one indicates the a perfect sphere, here all masses have sphercity value less than 1.

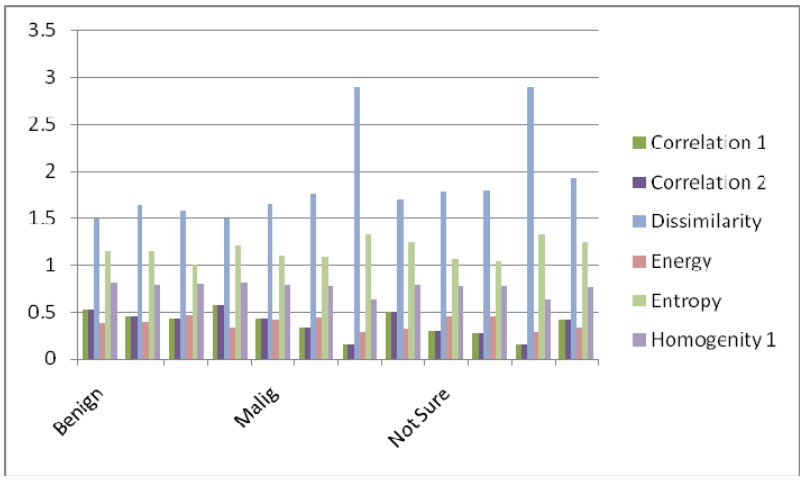

Fig 2. Feature values for different patients

Table 3 shows the Recognition rate of Texture algorithms

Feature Extraction Technique	**Value (in %)**	**Feature Extraction Technique**	**Value (in %)**
Auto correlation	100	Sum average	100
Contrast	95	Sum variance	100
Cluster prominence	86	Sum entropy	59
Cluster Shade	91	Solidity	91
Correlation1	100	Dissimilarity	95
Correlation2	100	Difference entropy	95
Homogeneity1	96	Nrl Entropy	95
Homogeneity 2	96	Sum of squares	100

Performance evaluation is done to identify the cases which correctly identified the feature values ranges of the malignant and benign areas. Recognition rate of feature extraction techniques is shown in table 3. The plateau region's categorization can be done as a future work.

IV. Conclusion

In this work different feature extraction techniques for DCEMRI on the basis of various approaches to texture analysis are described. The features are classified as structural, statistical, model-based and transform based features. The statistical features and the normalized radial length features and Intensity time kinetic feature values are calculated for a set of images. The images are able to categorize as benign, malignant or 'not sure' based on the feature values. Then the image is segmented in to malignant, benign and ' not sure' region and the segmented portions feature values are calculated separately. Some features shows a peculiarity that, either the malignant or benign regions feature values are greater. Based on the assessment, the validity of the procedure is checked.

References

[1] A.Rosenfeld and A. Kak, (1982). Digital Picture Processing, vol. 1, Academic Press.

[2] Cai, H., Peng, Y., Ou, C., Chen, M., & Li, L. (2014). Diagnosis of breast masses from dynamic contrast-enhanced and diffusion-weighted MR: a machine learning approach. PloS one, 9(1), e87387.

[3] Chou, Yi-Hong et al., (2001), Stepwise logistic regression analysis of tumor contour features for breast ultrasound diagnosis Ultrasound in Medicine and Biology, Volume 27 Issue 11 , 1493 - 1498.

[4] Chu A., Sehgal C.M.A.and Greenleaf J.F., (1990), Use of gray value distribution of run lengths for texture analysis, Pattern Recognition Letters, Volume 11(6), Pages 415-419

[5] G. Malu, Elizabeth Sherly, Sumod Mathew Koshy, J. (2015), An automated algorithm for lesion identification in dynamic contrast-enhanced MRI, International Journal of Computer Applications in Technology, Vol. 51, No. 1.

[6] Malu, G., Balakrishnan, K., & Bodhey, N. K. (2011, March). Area and Volume Calculation of Necrotic Tissue regions of heart using Interpolation. In Emerging Trends in Electrical and Computer Technology (ICETECT), 2011 International Conference on (pp. 728-730). IEEE.

[7] Galloway M. M. (1975), Texture analysis using gray level run lengths, Computer Graphics and Image Processing, Volume 4(2), Pages 172-179[14].

[8] A. Gebejes, & Huertas, R. (2013). Texture Characterization Based on Grey-Level Co-Occurrence Matrix

[9] Haralick R., Shanmugan K., Dinstain I. , Textural Features for Image Classification, (1973), IEEE Transactions on Systems, Man and Cybernetics, 3, 6, 610-622

[10] Yao, J., Chen, J., & Chow, C. (2009). Breast tumor analysis in dynamic contrast enhanced MRI using texture features and wavelet transform. IEEE Journal of selected topics in signal processing, 3(1), 94-100.

[11] Keyvanfard, F., Shoorehdeli, M. A., & Teshnehlab, M. (2011, June). Feature selection and classification of breast MRI lesions based on Multi classifier. In Artificial Intelligence and Signal Processing (AISP), 2011 International Symposium on (pp. 54-58). IEEE.

[12] Kilday et al. (1993), stepwise logistic regression analysis of tumor contour features for breast US diagnosis.

[13] Nixon, M., (2008), Feature extraction & image processing. Academic Press.

[14] Gibbs, P., & Turnbull, L. W. (2003). Textural analysis of contrast□enhanced MR images of the breast. Magnetic Resonance in Medicine, 50(1), 92-98.

[15] Pang, Y., Li, L., Hu, W., Peng, Y., Liu, L., & Shao, Y. (2012). Computerized segmentation and characterization of breast lesions in dynamic contrast-enhanced mr images using fuzzy c-means clustering and snake algorithm.Computational and mathematical methods in medicine, 2012.

[16] Selvarajah, S., & Kodituwakku, S. R. (2011). Analysis and comparison of texture features for content based image retrieval. International Journal of Latest Trends in Computing, 2(1).

[17] Pang, Z., Zhu, D., Chen, D., Li, L., & Shao, Y. (2015). A computer-aided diagnosis system for dynamic contrast-enhanced MR images based on level set segmentation and ReliefF feature selection. Computational and mathematical methods in medicine.