Chapter III
SEGMENTATION TECHNIQUES FOR SONOELASTOGRAPHIC BREAST IMAGES

3.1. Data Acquisition

Ultrasound elastography images of the breast were acquired using Siemens Acuson Antares ultrasound scanner, with a 7.3 MHz linear array transducer. When acquiring the elasticity images slight compression is required and higher levels of pressure should be avoided. The effect of breathing and heartbeat produces the required compression. Elasticity imaging provides a live dual image display of both the standard B-mode ultrasound image and the elastography image. The techniques have been implemented with 50 benign and 50 malignant breast images acquired from Ultrasound elastography.

3.2. Pre-Processing

Breast ultrasound elastogram pre-processing consists of speckle reduction and image enhancement. Speckle is similar to the multiplicative noise showed by a number of scatterers with stochastic stage on the ultrasound resolution cell beams [20].

Noise can be defined as the stochastic change of the ideal pixel measure during image acquisition. This implies that, at the instance a measured pixel value can be perceived, the ideal value has already been distorted by a stochastic constituent. US imaging system experiences speckle noise to certain magnitude.

The noise effect model is:

\[ g(x, y) = f(x, y) + \eta(x, y) \]  

where \( g(x, y) \) represents the observed image, \( f(x, y) \) represents the unidentified undistorted image, and \( \eta(x, y) \) serves as the noise constituent. Since noise is naturally a stochastic component in the imaging process, it is difficult to reconstruct the ideal pixel measure \( f \) from
the observed measure $g$. It is possible to envisage some comprehensive consequences of noise, if the statistical behavior of the noise is known.

Practically, the assumption that noise is additive Gaussian noise does not hold true. A better advanced model accepts that noise is made up of two parts; a multiplicative component $\eta_1$ dependent on the pixel measure and an additive part $\eta_2$:

$$\eta(x, y) = h(f(x, y)) \cdot \eta_1(x, y) + \eta_2(x, y). \quad (3.2)$$

It is a safe assumption that both $\eta_1$ and $\eta_2$ demonstrates a normal distribution. An important exception to this is speckle noise, a multiplicative noise with a Poisson distribution. The models in the equations 3.1 and 3.2 are dependent only on the ideal measures of $f(x, y)$. This is not comprehensive enough to accurately model the noise. It therefore becomes necessary to integrate the neighborhood pixels around $(x, y)$ rather than using $(x, y)$ only.

A speckle field can be described by its statistical moment like any other ergodic process. The moments of a speckle field can be found by considering the moments of the fundamental Gaussian distribution function by employing the moment generating function for the complex Gaussian density function [45]. Assuming $x_i$ denotes a point from a speckle pattern, they are linked to complex Gaussian random variables $\eta_i$ by $y|n_i|$. The vector

$$\eta = \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_N \\ n_1^* \\ \vdots \\ n_N^* \end{bmatrix}$$

has density function

$$f(\eta) = \frac{1}{\pi^N |R|} \exp \left( \frac{1}{2} (\eta - \bar{\eta})^T [R \ 0] - 1 (\eta - \bar{\eta}) \right) \quad (3.4)$$

where $E [(n - \bar{n})(n - \bar{n})^H] = R, \bar{\eta}$ is the mean of $\eta, \bar{n}$ is the mean of $n$, and $N$ is the dimension of $n$. The Nth moment $E [\Pi_{i=1}^N x_i] = E [\Pi_{i=1}^N n_i n_i^*]$ can be calculated using
\[
E \left[ \prod_{i=1}^{N} n_i n_i^* \right] = \frac{\partial^{2N}}{\partial v_1 \partial v_2 \ldots \partial v_{2N}} E[\exp(v^H \eta)] \bigg|_{v=0}
\]

\[
= \frac{\partial^{2N}}{\partial v_1 \partial v_2 \ldots \partial v_{2N}} \left\{ \begin{array}{c} 
\exp(v^H \eta) \\
\exp \left( \frac{1}{2} v^H S v^H \right) 
\end{array} \right\} \bigg|_{v=0} (3.5)
\]

where \( S = \begin{bmatrix} 0 & R \\ R^T & 0 \end{bmatrix} \)

A valuable notion in examining and processing UE images is that a set of “sticks”, a line section of continuous length and adaptable orientation, which can approximate definite image functions. Sticks filter has been found useful in automatic segmentation techniques in detecting tumor boundaries.

When the noise is hypothetical white, then the issue of noise removal is reduced to a simple linear projection operation. At each point, the detection statistic is produced by adding all the pixel intensities falling along a stick of one orientation and maximizing the sum over all possible stick rotations:

\[
\Lambda(x) = \max_i (\mu^{(i)})^T x. 
\]

(3.6)

Although speckle is neither additive, nor Gaussian or white, it has been proposed that the “sticks” method to be suitable for detection of linear components as well as for image enhancement.

The concept behind this idea is that images can be enhanced by projecting the image pixels onto a set of “sticks”. The algorithm works to enhance the image by plotting sequentially at each point the maximum measure found by adding up pixels on a stick traversing the point, and changing the predilection of the stick. The stick method as defined by Czerwiniski et al. was used for our pre-processing operations.
3.3. Feature Set Optimization using Statistical Measures for UE Image Segmentation

This work deals with tumor boundary extraction by multiple features using statistical methods: GLCM, LBP and Edge operators. This is achieved by using statistical moments and comparative position of neighboring pixels in an image, as well as functions that counter monotonic brightness variations with computational simplicity. This is simply transforming the input data into feature vectors which solves the problem of feature selection while defining the data with sheer accuracy. The redundant and non-significant features are removed and thereby segmented and evaluated.

3.3.1. Feature Extraction

Feature extraction is a vital phase in breast tumor detection and classification. An optimum number of feature set is usually used as an effective and efficient feature discriminator. This mainly reduces the space occupied by the redundant features and guides against computational complexities or “curse of dimensionality”. The “curse of dimensionality” proposes that the sampling density of the training data is too low to promise a meaningful estimation of a high dimensional classification function with the available finite number of training data. In most sophisticated classification techniques, like support vector machine (SVM) and artificial neural network (ANN), the feature vectors dimensions extremely affects the performance and execution time of the algorithm. Hence, feature extraction is a fundamental operation for CAD systems.

The US breast image features can be broadly be divided into four groups: texture, model-based, morphologic and descriptor features. The universal rule in selecting optimum feature sets includes these contemplations: optimality, independence, discriminating properties and reliability [55]. Though, the mere combination of the best performing feature
set doesn’t guarantee an effective system. The chief aim of feature extraction is to exploit optimally the discriminating operations of the feature sets.

3.3.1.1. Texture Features

Texture is said to be the characteristics constituting the spatial organization of the gray levels of the pixel in a pixel neighborhood. The singular objective of a texture analysis approach is to define different textures existing in an image. Scale has been recognized as one of the critical characteristics of texture description [73]. Harnessing these descriptive features that is correlated to various texture scales, makes it possible to identify and distinguish different textures existing in an image.

Texture is a crucial feature for unsupervised interpretation of breast images. The research on texture analysis has been on for several decades. In the last three decades the focus has been on the algorithms based chiefly on the first and second order statistics relating to the gray level measures of the image pixels, referred to as the spatial domain gray level co-occurrence matrix and neighboring gray level dependence matrix. A shift in the mid 80’s gave rise to the model based approaches such as simultaneous autoregressive and Markov random field, followed by wavelet approaches which gained prominence in the late 80’s.

Texture analysis plays a central role in breast image segmentation and classification. In region description, the textural feature provides a valuable insight about the region than the using intensity descriptors like the mean, maximum and minimum gray-levels of the image pixels. The spectral and statistical methods are basic in texture analysis [34].

The spectral approach refers to the frequency domain and is a vital means of analyzing texture. Here features are linked to the filter responses statistics. Research has shown that there is a close resemblance in the structure of the neural sensory fields in the human optical system and the tuned band-pass filter bank. The chief drive of the spectral
approach is the extension of feature extraction into the spatial frequency domain. The spatial frequency has various advantages for feature extraction. One of these is that filters are selective, meaning that only specific features are enhanced while suppressing others. The Laplacian of Gaussian filters exemplifies this as they are known to be good edge detectors, while for blob detection the low-pass Gaussian filters are generally useful. The spatial frequency in the spectral domain can represent the cyclicity construction of a texture unambiguously.

Spectral approaches generally employ the use of filters for image conversion from spatial domain into the frequency domain and the other way around. The filter bank mainly consist of ambit of collimate filters, with filters set at a specific spatial frequency, predilection or scale. Information can be compressed through crumpling a specific frequency crosswise or through crumpling all frequency in a specific direction. This procedure allows one-dimensional descriptors practicable for texture discrimination. Nevertheless, the ensuing description is over-finished, as it contains large content and as well redundant information.

A common feature inherent in the procedures is the notion of separating/clustering pixels using the prevalent eigenvector of $n \times n$ matrix deduced from the matched attractions between pixels. Here $n$ represents the image pixels measure. The level of similarity measured by one or more discriminative stimulus is derived by computing the affinity between pixels.

The problem associated with the spectral methods, is that filters selection is heuristic and task dependent. Ensuing feature description is excessively completed and possibly giving rise to mingling of both important and orthogonal features.

In the statistical approach the image signal statistics is mainly collected from the spatial domain and used as a feature descriptor. The commonly used statistics include moments, identification histogram and the gray-level co-occurrence matrixes. In texture analysis the first and second-order statistics are normally exploited. The first-order statistics
like the mean, moment, standard deviation and moments of the histogram generally accounts for the attributes of the specific pixels. The second order statistics are concerned with the inter-dependency in the spatial domain.

One of the basic ways of texture description is by employing the statistical moments of the intensity histogram of an image. The drawback in using only the histogram in computation is that only the values with intensities distribution will be taken into account, but not accounting for the pixels relative position with one another in the image texture. Employing a statistical technique like the co-occurrence matrix ensures that vital information about the neighboring pixels relative position in an image is harnessed.

GLCM approximates the image attributes linked with the second-order statistics cogitates the association between pixels or groups of pixels (commonly two). Using a one-dimensional histogram may not be realistic in depicting texture features because of its spatial property. Therefore, GLCM (Two dimensional) is broadly employed in texture analysis and is briefly discussed below.

### 3.3.1.1.1. Gray Level Co-Occurrence Matrix (GLCM)

The gray level co-occurrence matrix, a two-dimensional (2D) matrix is extensively used in texture analysis to capture the spatial reliance of the gray-level measures required for texture perception.

The Gray Level Co-occurrence Matrix (GLCM), which is a square matrix, is capable of divulging certain properties about the spatial distribution of the gray-levels within the texture image. The GLCM is a statistical texture measure which gathers information about pixel sets. It is therefore a second order statistic which tabulates the frequencies or pixel brightness values in an image.
The matrix is a construct of a distance $d$ and direction of $\theta$ (given as $0^\circ$, $45^\circ$, $90^\circ$ and $135^\circ$). Textural features are usually extracted using GLCM contrast, correlation, energy, entropy, local homogeneity, and shade.

The GLCM describes the specific relationship between the references pixel with intensity measure $i$ to the neighborhood pixel with the intensity measure known as $j$. Hence, the relative distance $d$ between $i$ and $j$ is the number of occurrence of each element $(i, j)$ of the matrix. The spatial association concerning two neighboring pixels can be represented in diverse means with diverse origins and angles, with the immediate neighboring pixel to right as the default pixel. The contiguity can be described to occur in each of the four directions (vertical, horizontal, right and left diagonal as shown in Figure 3.1.

Precisely, for a given image $I$ of size $K \times K$, the elements of a $G \times G$ gray-level co-occurrence matrix $M_{CO}$ for a displacement vector $d = (d_x, d_y)$ is defined as

$$M_{CO} = \sum_{x=1}^{k} \sum_{y=1}^{k} \begin{cases} 1, & \text{if } I(x, y) = i \text{ and } I(x + d_x, y + d_y) = j \\ 0, & \text{otherwise} \end{cases} \quad (3.7)$$

Figure 3.1 illustrates the details of the process to generate three symmetrical co-occurrence matrices considering a $3 \times 3$ image represented with four gray tine values from 0 to 3. For the purpose of this research we considered one neighboring pixel ($d=1$) along one possible direction as $[-1 \ 0]$ for $90^\circ$.

![Figure 3.1. Co-occurrence matrix direction for extracting texture features](image_url)
Each element of the GLCM is the number of times that two pixels with gray tone i and j are neighborhood in distance d and direction θ. For d=0 co-occurrence matrix, there are two occurrences of the pixel intensity value 1 and pixel intensity value 3 adjacent to each other in the input image. Also, the occurrence of pixel intensity value 3 and pixel intensity value 1 adjacent to each other is 2 times. Hence, these matrices are symmetric in nature and the co-occurring pairs obtained by choosing θ equal to 0° will resemble what is observed by choosing θ equal to 180°. This concept extends to 45°, 90° and 135° as well. With all these considerations, the GLCM matrix is calculated for each of the four possible angles.

In order to extend the concept of co-occurrence matrices to n-dimensional Euclidean space, a mathematical model for the above concept is required. Here the universal set is set as \( Z^n \) and \( Z^n \) is the Cartesian product of Z taken n times itself, where Z is the set of all integers. A point (or pixel in \( Z^n \)) X in \( Z^n \) is an n-tuple of the form \( X=(x_1, x_2, ..., x_n) \) where \( x_i \in Z \) \( \forall i = 1,2,3,...,n \). An image I is a function from a subset of \( Z^n \) to Z. That is \( f: I \rightarrow Z \) where \( I \subset Z^n \). If \( X \in I \), then X is assigned an integer Y such that \( Y=f(X) \). Y is called the intensity of the pixel X. The image is called a gray scale image in the n-dimensional space \( Z^n \).

The generalized Co-occurrence matrix for any gray scale image I defined in \( Z^n \) is obtained. The gray level co-occurrence matrix is defined to be a square matrix \( G_d \) of size N where, N is the total number of gray levels in the image. The (i, j)th entry of \( G_d \) represents the number of times a pixel X with intensity value I is separated from a pixel Y with intensity value j at a particular distance k in a particular direction d. The distance k is a non-negative integer and the direction d is specified by \( d=(d_1, d_2, d_3, ..., d_n) \), where \( d_i \in \{0, k, -k\} \) \( \forall i = 1,2,3,... \)

Consider a gray scale image Z, the co-occurrence matrix \( G_d \) for this image such that \( G_{-d} = G_d \). It can be seen that \( X+d=Y \), so that \( G_{-d} = G_d \) where \( G_d^t \) is the transpose of \( G_d \). Hence \( G_d^t = G_d \).
$G_{-d}$ is a symmetric matrix. Since $G_{-d} = G_d^\top$, it can be said that $G_d$ and $G_{-d}$ are dependent (or not independent). Therefore the direction $d$ and $-d$ are called dependent or not independent.

The theorem states: If $X \in \mathbb{Z}^n$, the number of independent directions from $X$ in $\mathbb{Z}^n$ is $\frac{3^n - 1}{2}$.

These are symmetric matrices hence evaluation of either upper or lower triangle serves the purpose. To normalize the frequency the value in each cell can be divided by the total number of possible pixel pairs. Other vital concerns in determining the co-occurrence texture features is the level of quantization and the components of the neighboring co-occurrence matrix which are associated greatly since they are values of analogous image characters.

The selection of the radius $d$ is a very important aspect in any research that employs GLCM. Several researches employed $d$ values from the range of 1 to 10. In a fine texture analysis using a big translational value will give rise to a GLCM that fails to capture the textural information elaborately. From the research reviewed, it can be concluded that the optimal classification accuracies occur with $d$ set at 1, 2, 4, 8, and with the most beneficial result at $d = 1$ and 2. This is justifiable, as correlation between pixels is more likely to occur with the neighboring pixel than a pixel located far off. Furthermore, translational value equivalent to the size of the texture component enhances the classification potentials. For the purpose of this research, the GLCM features obtained from distance 1 is used. To get an accurate segmentation on a medical image minimal distance is to be taken.

Also the selection of the angle $\theta$ is to be done for obtaining GLCM and further for the calculations of the features. In each pixel there are eight adjacent pixels permitting for eight options for $\theta$, which are $0^\circ$, $45^\circ$, $90^\circ$, $135^\circ$, $180^\circ$, $225^\circ$, $270^\circ$ or $315^\circ$. Nonetheless, considering the fundamental concepts of GLCM, the co-occurrence of pairs found by choosing $\theta$ equal to $0^\circ$ would be comparable to those found by choosing $\theta$ equal to $180^\circ$. This notion stretches to $45^\circ$, $90$ and $135^\circ$ as well.
Therefore, it is narrowed down to four options for the value of $\theta$. At times, isotropic
GLCM can be obtained by integration at all angles if the image is isotropic and directional
information is not needed. However, this research was conducted using an angle of 90°.

Next comes the selection of the quantized gray levels $G$. The maximum gray value of
the pixels determines the GLCM dimensions. In GLCM computation the amount of gray
levels is very essential. An increase in levels translates to accuracy in extracted textural
information but with increase in computational costs. Computational complexities are highly
correlated to the number of gray levels and are relative to $\theta$ [29]. Hence, if the value of $G$ is
predetermined, a GLCM is needed for every singular pair of $d$ and $\theta$.

The lower left triangular matrix of the GLCM is most times a mirror image of the
upper right triangular matrix, with the slope containing mostly even numbers. Several GLCM
parameters are associated to definite first-order statistical models. For example, contrast
would be the same as pixel pair recurrence rate, and the variance might be taken as the spatial
frequency detection etc. Conventionally, the number of gray levels $G$ determines the GLCM
dimension and keeps the co-occurrence. To analyze the texture features, iteration through the
entire matrix is performed by applying the selected statistics. The textural features are
grounded on statistics which encapsulate the relative frequency distribution that describes the
frequency of a specific gray tones appearance in a particular spatial relationship with another
gray tone in the image.

The texture analysis technique employed in this study was constructed on a co-
occcurrence matrix of the distribution of gray-level values. Each entry $(i, j)$ in the GLCM
relates to the number of occurrences of the pair of gray-levels $i$ and $j$ that are $d$ at a distance in
the input image and lengthwise direction $\theta$. Furthermore, potential computational
complexities were circumvented by using matrices obtained in one direction ($\theta = 90^\circ$). A
number of parameters can be computed from the GLCM in classifying benign tumors from
malignant ones in breast US images. In this study the 18 parameters computed are autocorrelation, contrast, correlation, cluster prominence, cluster shade, dissimilarity, energy, entropy, homogeneity, maximum probability, variance, sum average, sum variance, sum entropy, difference variance, difference entropy, information measure of correlation and inverse difference normalized.

The texture-feature parametric imaging used a sliding overlapping window with $m$-by-$m$ pixel in the gray-level image to evaluate each local texture feature. A $3 \times 3$ window was prompted diagonally the whole matrix of the gray level image sequentially one pixel at a time and assigned the local texture feature as the current pixel located in middle of the window. This process produced a texture feature parametric image as a function of texture feature parameter measures.

3.3.1.1.2. Local Binary Pattern (LBP)

LBP operator is a combination of statistical and structural texture analysis features. It is generally employed in statistical gray invariant 2-D texture analysis. The operator starts by labelling the image pixels through thresholding of the neighborhood (3 x 3) of respective pixel with the center measure and regarding the outcome of this thresholding as a binary number.

After labeling the pixels with the matching LBP codes, histogram of the labels are calculated and employed as a texture descriptor. The most appealing properties of LBP operators are permissiveness against monotonic elucidate changes and computational ease. The basic LBP operator is impracticable for central feature of an enormous scale structures. Hence, the need to lengthen it to expedite the analysis of textures having numerous scales bycoalescing neighborhoods with dissimilar dimensions.
Number of years after it was originally proposed by Ojala et al. [85], the local binary pattern operator was introduced in a more broad reviewed system. Contrary to the rudimentary LBP employing 8 pixels in a 3 x 3 block, the new broad conceptualization of the operator has no limitations to the dimension of the neighborhood or to the number of sampling point that can be used.

Contemplate a monotonous image $I(x, y)$ and let $g_c$ represent the gray level of a discretionel pixel $(x, y)$, meaning $g_c = I(x, y)$.

Furthermore, let $g_p$ refer the gray value of a sampling point in a uniformly spaced circular neighborhood of $P$ sampling points and radius $R$ around $(x, y)$:

$$g_p = I(x_p, y_p), \quad p = 0, ..., P - 1$$ (3.8)

$$x_p = x + R \cos(2\pi p/P)$$ (3.9)

$$y_p = y - R \sin(2\pi p/P)$$ (3.10)

Taking that an image’s local texture $I(x, y)$ is specified by the joint distribution of gray values: $P + 1$ ($P > 0$) pixels:

$$T = t(g_c, g_0, ..., g_{P-1})$$ (3.11)

Without losing any information, the center pixel value can then be subtracted from the neighborhood:

$$T = t(g_c, g_0 - g_c, g_1 - g_c, ..., g_{P-1} - g_c)$$ (3.12)

Moving further the joint distribution is estimated by taking the center pixel to be statistically autonomous of the differences, which allows for factorization of the distribution:

$$T \approx t(g_c) t(g_0 - g_c, g_1 - g_c, ..., g_{P-1} - g_c)$$ (3.13)

The first factor $t(g_c)$ represents the intensity distribution over $I(x, y)$. Considering the analysis of local textural patterns, this contains no practicable information. Rather the joint distribution of differences

$$T \approx t(g_0 - g_c, g_1 - g_c, ..., g_{P-1} - g_c)$$ (3.14)
is more practicable to use in modelling the local texture. Nonetheless, dependable estimation of this multidimensional distribution from the image data can be problematic. Applying vector quantization is one solution to this problem proposed by Ojala et al. in 2002. They used a codebook of 384 code words with learning vector quantization to achieve reduction in dimensionality in high dimensional feature space. The 384 code words indicators are matched to the 384 bins found in the histogram.

However, the learning vector quantization based techniques still has some challenges that makes it problematic in reality. One, the dissimilarities \( g_p - g_c \) are invariant to alterations of the image’s mean gray value but is not affected by other changes in gray levels. Two, for it to be effective in texture classification the codebook must be aimed like other texton-based techniques. To counter these problems, it is best to consider only the signs of the differences.

\[
T \approx t(s(g_0 - g_c), s(g_1 - g_c), \ldots, s(g_{P-1} - g_c)),
\]

(3.15)

\( s(z) \) is known as the thresholding function

\[
s(z) = \begin{cases} 
1, & z \geq 0 \\
0, & z < 0.
\end{cases}
\]

(3.16)

The broad local binary operator results from this joint distribution. In the case of the fundamental LBP, it is derived by summation of the threshold differences weighted by powers of two. This \( \text{LBP}_{P,R} \) operator is specified as

\[
\text{LBP}_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c)2^p
\]

(3.17)

Practically, equation 3.10 implies that the signs of the differences in a neighborhood are inferred as a \( P \)-bit binary number, giving rise to \( 2^P \) distinctive measures for the LBP code. The local gray scale distribution can be more or less defined with a \( 2^P \) bin discrete distribution of LBP codes:

\[
T \approx t(\text{LBP}_{P,R}(x_c, y_c))
\]

(3.18)
In computing the LBP$_{P,R}$ feature vector for a given $N \times M$ image sample $(x_c \in \{0, \ldots, N-1\}, y_c \in \{0, \ldots, M-1\})$, the central part is only given consideration because it is impracticable to use a large neighborhood on the borders. The LBP code is computed for specific pixel in the trimmed parts of the image, and the distribution of the codes is used as a feature vector, indicated by $S$:

$$S = t(\text{LBP}_{P,R}(x,y))$$  \hfill (3.19)

where $x \in \{[R], \ldots, N-1-[R]\}, y \in \{[R], \ldots, M-1-[R]\}$.

The Rotational invariant LBP feature is obtained from the LBP feature vector.

Taking $U_p(n,r)$ to represent a particular uniform LBP pattern, the duo $(n,r)$ defines a uniform pattern in such a way that $n$ is the number of 1-bits in the pattern and $r$ is the rotation of the pattern.

Assuming the neighborhood has $P$ sampling points then $n$ derives its values from 0 to $P + 1$, where $n = P + 1$ is the special label noting all the non-uniform patterns. Moreover, when $1 \leq n \leq P - 1$, the patterns rotation will be in the range $0 \leq r \leq P - 1$.

Taking $I^\alpha(x,y)$ represent the image’s rotation $I(x, y)$ by $\alpha$ degrees. With this rotation, point $(x, y)$ is rotated to location $(x', y')$. Neighborhood with circular sampling points $I(x, y)$ and $I^\alpha(x', y')$ also rotate by $\alpha^\circ$.

Thus, if the rotation are limited to integers multiples of the angle between two sampling points; $\alpha = a \frac{360^\circ}{P}$, $a = 0, 1, \ldots, P - 1$, the sampling neighborhood is then rotated by exactly $\alpha$ discrete steps. Hence, the uniform pattern $U_p(n,r)$ at point $(x, y)$ is substituted by uniform pattern $U_p(n,r + a \ mod \ P)$ at point $(x', y')$ of the rotated image.

Taking into consideration, the original rotation invariant LBPs, the rotation invariant features can be derived. As stated earlier, the textured input image’s rotation induces the LBP patterns to transform into a different location and rotate about their origin. Calculating the histogram of LBP codes normalizes for transformation, and normalization for rotation is
accomplished by rotation invariant mapping. With this mapping, every LBP binary code is rotated circularly into its lowest value.

\[ LBP_{F,R}^{ri} = \min_i \, ROR \left( LBP_{F,R,i} \right) \]  

(3.20)

With \( ROR \, (x, i) \) representing the circular bitwise right rotation of bit sequence \( x \) by \( i \) steps. If the sampling artifacts are excluded, then the histogram of \( LBP_{F,R}^{ri} \) codes is invariant only to rotations of input images by angles \( \omega \). Nonetheless, from experiment this descriptor is shown to be very robust to in-plane rotations of images by any angle.

Above the description of the rotational invariance LBP features and how they are extracted from the 50 benign and 50 malignant UE images has been given. The feature vector needed for the next stage of processing is created from the LBP features. The UE images are divided into \( N \times N \) overlapping sub-images. The LBP features are extracted from each sub-image to create the feature vector needed for the next phase of the unsupervised segmentation. The feature vectors dataset of all the images are utilized in the next phase of feature selection.

### 3.3.1.2. Edge Features

An Edge in an image is a significant local change in the image intensity, usually associated with a discontinuity in either the image intensity or the first derivative of the image intensity. Discontinuities in the image intensity can be either Step edges, where the image intensity abruptly changes from one value on one side of the discontinuity to a different value on the opposite side, or Line Edges, where the image intensity abruptly changes value but then returns to the starting value within some short distance. However, Step and Line edges are rare in real images. Because of low frequency components or the smoothing introduced by most sensing devices, sharp discontinuities rarely exist in real
signals. Step edges become Ramp Edges and Line Edges become Roof edges, where intensity changes are not instantaneous but occur over a finite distance [91].

Changes or discontinuities in image attribute such as luminance or tristimulus value are the fundamentally important primitive characteristics of an image because they often provide an indication of the physical extent of objects within the image.

There four types of one-dimensional uninterrupted domain edges as shown in Figure 3.2. The ramp edge is patterned in form of a ramp, increase in the ramp corresponds to increase in image amplitude from low to high level, or in opposite direction. The edge is defined by these characteristics; height, angle of the slope and the orientation of the midpoint of the slope. An edge is said to exist if the measure of the edge height is larger than the indicated value. Idealistically, an edge detector should be able to create an edge reading confined to a single pixel positioned at the center of the slope. When the angle of the slope of a ramp edge is equal to 90° the resulting edge is known as a step edge. Step edges exist only for artificially produced images like test patterns. In line edges, when the line reaches its limit, as the width of the line draws close to zero, the resulting amplitude discontinuity is known as roof edge.

![Types of Edges](image)

**Figure 3.2.** Types of Edges (a) Step Edge (b) Ramp Edge (c) Line Edge (d) Roof Edge
Although point and line detection certainly are important in any discussion on segmentation, edge detection is by far the most common approach for detecting meaningful discontinuities in gray level images. Edges were introduced as a set of connected points that lie on the boundary between two regions.

Zernike moments, a type of moment function, are the mapping of an image onto a set of complex Zernike polynomials. As these Zernike polynomials are orthogonal to each other, Zernike moments can represent the properties of an image with no redundancy or overlap of information between the moments. Due to these characteristics, Zernike moments have been utilized as feature sets in applications such as edge detection, content-based image retrieval, and other image analysis systems [62].

The computation of Zernike moments from an input image consists of three steps: computation of radial polynomials, computation of Zernike basis functions, and computation of Zernike moments by projecting the image onto the basic functions. In this section, the procedure for the computation of Zernike moments is described. The procedure for obtaining Zernike moments from an input image begins with the computation of Zernike radial polynomials. The real-valued 1-D radial polynomial $R_{nm}(\rho)$ is defined as

$$R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} c(n, m, s) \rho^{n-2s}$$ (3.21)

where

$$c(n, m, s) = (-1)^s \frac{(n-s)!}{s!(n + |m|)/2 - s)!((n - |m|)/2 - s)!}$$

But the refined model proposed by Boland[19] was used in edge detection for the purpose of this research. According to Boland, Zernike moment can be defined over the unit circle, and involves two steps to convert an image’s rectangular region to a unit circle for the calculation of Zernike moments. In the first step the center of the data image was calculated and used to denote the pixel coordinate system’s mid-point. In stage two, involves dividing the x and y coordinates. And using only the pixels that falls within the resultant normalized
image, \( f(x, y) \) in the later calculations. The Zernike moments, \( Z_{nl} \), for an image were then computed using

\[
Z_{nl} = \frac{n+1}{n} \sum_x \sum_y V_{nl}^* (x, y) f(x, y)
\]

(3.22)

where \( x^2 + y^2 \leq 1, 0 \leq l \leq n, n - l \) is even, \( f(x, y) \) depicts the intensity measure of the normalized image and \( V_{nl}^* \) representing composite conjugate of a Zernike polynomial with degree \( n \) and angular dependence \( l \)

\[
V_{nl}^* (x, y) = \sum_{m=0}^{n-1} (-1)^m \frac{(n-m)!}{m!(n-2m+1)!} \left( \frac{x^2 - y^2}{2} \right)^{n-m} e^{i\theta}
\]

(3.23)

where \( 0 \leq l \leq n, n - l \) is even, \( \theta = \tan^{-1} \left( \frac{y}{x} \right) \) and \( i = \sqrt{-1} \).

### 3.3.2. Feature Selection

Feature selection is very vital in applications involving pattern recognition. In most medical diagnostic applications, there is a critical need to assess the potency of various feature groupings and use the most optimal for classification. Using only the optimal subset of features used in classification reduces the computational time immensely.

Feature selection simply means to select a subset of \( x \) features from a bigger set of \( y \) features with the aim of optimizing the measure of a benchmark function \( J \) over all the subset of the scope of \( x \). The many different feature selection algorithms have been reviewed in the section dealing with literature review. The sequential forward selection (SFS) and sequential backward selection (SBS) are two of the most widely used feature selection technique. The SFS techniques starts by selecting the most beneficial individual feature and subsequently adds more feature one at a time in compounding with the selected features till the benchmark function \( J \) is maximized. The SBS on the other hand sets out with all input feature and the SBS method starts with all input features and sequentially removes features.
The highpoint of both the SFS and SBS is their computational ease. Through dynamic adjusting of the number of forward and backpedalling steps, floating sequential search technique have been shown to outperform both the SBS and SFS algorithms.

The Sequential Forward Floating Selection (SFFS) algorithm starts the exploration without any feature set and employs the SFS algorithm to select features one after another to the selected feature subset. Each time a new feature is selected and added to the feature subset, there is a backtracking by the algorithm by means of the SBS deleting one feature at time and locating and replacing it with a better subset. The exploration comes to an end when the dimension of the present feature set exceeds that of feature $x$ desired.

SFFS is famous and extensively used because of lack of computational complexities and cost-effectiveness with respect to the resulting optimized solution.

Before depicting the representative SFFS algorithms, the following definition is highlighted.

Let $X_k = \{x_i: 1 \leq i \leq k, x_i \in Y\}$ be the set of $k$ features from the set $Y=\{y_i: 1 \leq i \leq D\}$ of $D$ obtainable features. The measure $J(y_i)$ of the feature selection standard is valid if only the ith feature $y_i, i = 1, 2, \ldots, D$, is used and will be called the single significance $S_0 (y_i)$ of the feature.

The significance $S_{k-1}(x_j)$ of the feature $x_j, j = 1, 2, \ldots, k$ in the set $X_k$ is defined by

$$S_{k-1}(x_j) = J(X_k) - J(X_k - x_j) \quad (3.24)$$

The significance $S_{k+1}(f_j)$ of the feature $f_j$ from the set $Y - X_k$

$$Y - X_k = \{f_i: i = 1, 2, \ldots, D - k, f_i \in Y, f_i \neq x_i \text{ for all } x_i \in X_k\}$$

With respect to the set $X_k$ is defined by

$$S_{k+1}(f_j) = J(X_k + f_j) - J(X_k) \quad (3.25)$$

For $k = 1$ the term feature significance in the set coincides with the term of individual significance.
Let say that the feature \( x_j \) from the set \( X_k \) is

(a) The most significant feature in the set \( X_k \)

\[
S_{k-1}(x_j) = \max_{1 \leq i \leq k} S_{k-1}(x_i) \Rightarrow J(X_k - x_j) = \min_{1 \leq i \leq k} J(X_k - x_i) \quad (3.26)
\]

(b) The least significant (worst) feature in the set \( X_k \) if

\[
S_{k-1}(x_j) = \min_{1 \leq i \leq k} S_{k-1}(x_i) \Rightarrow J(X_k - x_j) = \max_{1 \leq i \leq k} J(X_k - x_i) \quad (3.27)
\]

Let say that the feature \( f_i \) from the set \( Y - X_k \) is

(a) The most significant (best) feature with respect to the set \( X_k \) if

\[
S_{k+1}(f_i) = \max_{1 \leq i \leq D-k} J(X_k + f_i) \quad (3.28)
\]

(b) The least significant (worst) feature with respect to the set \( X_k \) if

\[
S_{k+1}(f_i) = \min_{1 \leq i \leq D-k} S_{k+1}(f_i) \Rightarrow J(X_k + f_i) = \min_{1 \leq i \leq D-k} J(X_k + f_i) \quad (3.29)
\]

**Sequential Forward Floating Selection Algorithm**

Suppose \( k \) features have already been selected from the complete set of measurements \( Y = \{ y_j \mid j = 1, 2, \ldots, D \} \) to form set \( X_k \) with the corresponding criterion function \( J(X_k) \). Furthermore, the values of \( J(X_i) \) for all preceding subsets of size \( i = 1, 2, \ldots, k - 1 \) are known and stored.

**Step 1 (Inclusion)**

Using the basic SFS method, select feature \( x_{k+1} \) from the set of available measurements, \( Y - X_j \), to form feature set \( X_{k+1} \), i.e., the most significant feature \( x_{k+1} \) with respect to the set \( X_k \) is added to \( X_k \). Therefore \( X_{k+1} = X_k + x_{k+1} \).

**Step 2 (Conditional exclusion).**

Find the least significant feature in the set \( X_{k+1} \).

If \( x_{k+1} \) is the least significant feature in the set \( X_{k+1} \), i.e., \( J(X_{k+1} - x_{k+1}) \geq J(X_{k+1} - x_j) \), \( \forall j = 1, 2, \ldots, k \), then the set \( k = k + 1 \) and return to Step 1, but if \( x_r, 1 \leq r \leq k \)
$k$, is the least significant feature in the set $X_{k+1}$ i.e. $J(X_{k+1} - x_r) > J(K_k)$, then exclude $x_r$ from $X_{k+1}$ to form a new feature set $X'_k$ i.e. $X_k = X_{k+1} - x_r$.

Note that now $J(X_j) > J(X_k)$. If $k = 2$, then set $X_k = X'_k$ and $J(X_k) = J(X'_k)$ and return to Step 1, else go to Step 3.

**Step 3 (Continuation of conditional exclusion).**

Find the least significant feature $x_s$ in the set $X'_k$.

If $J(X'_k - x_s) \leq J(X_{k-1})$ then set $X_k = X'_k, J(X_k) = J(X'_k)$ and return to Step 1. If $J(K'_k - x_s) > J(X_{k-1})$ then exclude $x_s$ from $X'_k$ to form a newly reduced set $X'_{k-1}$ i.e. $X'_{k-1} = X'_k - x_s$. Set $k = k - 1$. Now if $k = 2$, then set $X_k = X'_k$ and $J(X_k) = J(X'_k)$ and return to Step 1, else repeat Step 3.

The algorithm is modified by setting $k = 0$ and $X_0 = \emptyset$, and the SFS method is used until a feature set of number 2 is achieved. Then the algorithm continues with Step 1.

### 3.3.3. K-means Clustering

The aim of the K-means clustering algorithm is to split an image into $K$ segments (using $K - 1$ thresholds), reducing the total within-segment variance. It is highly recommended that the variable $K$ is set before running the algorithm.

The within segment variance $\sigma^2_w$ is specified by

$$
\sigma^2_w = \sum_{i=0}^{k-1} h_i \sigma^2_i,
$$

where $h_i = \sum_v (s_i h(v))$ denotes the probability that a random pixel belongs to segment $i$ (containing the gray values in the range $S_i$), $\sigma^2_i = \sum_{v \in S_i} (v - \mu_i)^2 h(v)$ is the variance of gray values of segment $i$ and $\mu_i = \sum_{v \in S_i} vh(v)$ is the mean gray value in segment $i$.

Initialization: distribute the $K - 1$ threshold over the histogram. (For example in such a way that the gray value range is divided into $K$ pieces of equal length.). Image is then
segmented in agreement to the threshold set. For each segment, the “cluster center” is calculated, i.e., the value midway between the two thresholds that make up the segment.

1. For each segment, the mean pixel value $\mu_i$ is calculated.

2. The cluster centers is reset to the calculated values $\mu_i$.

3. The threshold is reset to be midway between the cluster centers, and segment the image.

4. Step 2 is returned to. Iteration is performed until the cluster centers stops to move (or do not move significantly).

Though this algorithm minimizes variance does not require any variance to be calculated explicitly.

3.3.4. Morphological Processing

Morphological operations are used in pre-processing or post processing (thinning, filtering and pruning) stages of breast image processing and mainly applied on a gray scale image.

There are two main morphological operations: dilation and erosion [44]. Dilation involves object expansion, possibly filling the small holes and linking disjointed objects. Erosion primarily shrinks the object through eroding of their boundaries. Both morphological operations can be modified for a specific task by setting out with the right structuring elements, which influences precisely the dilation or eroding of the object.

The dilation method is achieved by resting the structuring element $B$ on the image $A$ and gliding it across the image in a way resembling convolution. The operation can be described in a sequence of steps:
1. When the origin of the structuring element corresponds with a 'white' pixel in the image, giving rise to no changes; then its best to graduate to the next pixel.

2. However, when the origin of the structuring element corresponds with a 'black' in the image, and then set to black the entire pixel from the image enclosed by the structuring element.

Representation: \( A \oplus B \)

The erosion method resembles dilation, nonetheless the pixels are turned to 'white', not 'black'. Like in the previous operation, the structuring element is glided throughout the image and then employing the following steps:

1. When the origin of the structuring element corresponds with a 'white' pixel in the image, giving rise to no changes; graduate to the next pixel.

2. When the origin of the structuring element corresponds with a 'black' pixel in the image, and perhaps with a minimum of one of the 'black' pixels in the structuring element falling over a white pixel in the image, the 'black' pixel in the image is changed from 'black' to a 'white'.

Representation: \( A \ominus B \)

The two elementary operations, dilation and erosion, can be compounded into more complex arrangements. The most beneficial of these for morphological filtering are known as opening and closing [124]. Opening simply comprises of an erosion process closely followed by a dilation, and can possibly be used in removing pixels in regions where the structuring elements cannot be contained because of the small pixel size. In essence what the structuring element does is probe the image for small objects to be filtered out of the image.

Representation: \( A \circ B = (A \Theta B) \oplus B \)

Closing comprises of a dilation that is followed by erosion process and is used for filling of holes and small gaps.
**Boundary extraction**

The boundary of a set $A$, denoted by $\beta(A)$, is achieved by initially eroding $A$ by $B$ and then executing the set differences between $A$ and its erosion. That is, $\beta(A) = A - (A \Theta B)$, where $B$ is a suitable structuring element, ‘−’ is the difference operation on sets.

**Region filling**

Next a simple algorithm for region filling is developed and is grounded on set dilations, complementation, and intersections.

Starting with a point $p$ inside the boundary, the goal is to fill the entire region with ‘black’. Assuming that all non-boundary (background) points are considered ‘white’, assign a value of ‘black’ to $p$. The subsequently fill the region with ‘black’:

$$X_k = (X_{k-1} \oplus B) \cap A^C, \quad k = 1, 2, 3, ...$$

Where $X_0 = p$,

- $B$ is the corresponding structuring element
- $\cap$ is the intersection operator
- $A^C$ is the complement of set $A$

The algorithm terminates at iteration step $k$ if $X_k = X_{k-1}$. The set union of $X_k$ and $A$ contains the filled set and its boundary.

The aforementioned stages make up the morphological processing of the 100 breast images each of both benign and malignant tumors. The foundational stage comprises of eliminating the “small” objects from the breast images in order to clean up the background texture. To trail the conversion regions closely, edge smoothing is executed employing a progressive process of dilation followed intimately by erosion using a square mask of size $3 \times 3$. Lastly, disjunct background pixels are linked by filling the holes in the background texture. After this process, the image boundary is extracted which can be used for further classification.
3.3.5. Summary

This chapter introduced some new techniques for UE breast image segmentation. It gives a detailed step by step analysis of the distinctive procedures used in the acquisition, pre-processing, feature extraction, selection and finally the Segmentation.

Firstly, it shows the type of images used in this work and the pre-processing technique using stick and then feature extraction from GLCM, LBP and detecting edge features using Zernike moments. To counter the curse of dimensionality, optimal features were selected using SFFS and Segmentation is achieved using K-means algorithm.

The UE breast image pre-processing, segmentation and the feature extraction are the vital steps in the process because the accurate breast tumor classification will greatly depend on it. That has been dealt with thoughtfully and explicitly, in extraction and selection of optimal set of features for classification.

In the end, to solve a multi class problem the selected subset of optimal features were evaluated for their performance to validate the sub set that fits the data image and provide statistically meaningful result.

3.4 Multi-scale Texture Analysis for UE Image Segmentation

The wavelet transform a mathematical instrument and is found useful in great number of different applications, one area is found vital central to this research is in signal and image analysis as stated by Mallat [75]. Employing a wavelet transform is possible to split data into different scaling elements and then analyze each element with a resolution matched to its scale. There are various families of wavelet, for instance functions appropriate for continuous transformations and set of orthonormal function having compact support, which describes a vital set of discrete wavelet transforms, spline wavelets and various others.
In this work the family of orthonormal wavelets with compact support defined by Ingrid Daubechies is considered. This family of functions describes an orthonormal origin for the space of square integrable functions

\[ L^2(\mathbb{R}) = \{ f: \mathbb{R} \to \mathbb{C}: \int_{\mathbb{R}} |f(x)|^2 < \infty \} \]  

(3.31)

The Daubechies wavelet construct is found on a particular framework known as the multiresolution analysis, which is a group of attributes required for the construction of the foundation of \( L^2(\mathbb{R}) \), and also for defining this wavelet family, as well the scales relationships, and the attainment of the wavelet transform.

In this section the Multiresolution analysis is described using Ingrid Daubechies definitions. Nonetheless here the idea of filter construction is left out. For the purpose of this research the filters considered are taken as the ones in the list given by Ingrid Daubechies. The aim rather is to show how the discrete wavelet transform can be deduced from the scaling relations described by multiresolution framework.

### 3.4.1. Multiresolution Discrete Wavelet Transform

According to Ingrid Daubechies, a Multi Resolution Analysis (MRA) is represented by a family of subspaces \( V_j \in L^2(\mathbb{R}) \) that fulfills the following attributes:

**Condition I Monotonicity**

The sequence or function is consistently increasing, \( V_j \subset V_{j+1} \) for all \( j \in \mathbb{Z} \).

**Condition II Existence of the Scaling Function**

A function \( \varphi \in V_0 \) must exist, such that the set \( \{ \varphi(\cdot - k): k \in \mathbb{Z} \} \) is an orthonormal basis for \( V_0 \).

**Condition III Dilation Property**

For each \( j, f(x) \in V_0 \) if and only if \( f(2^j x) \in V_j \).

**Condition IV Trivial Intersection Property**
\[ \bigcap_{j \in \mathbb{Z}} V_j = \{0\}. \]

**Condition V Density**

\[ \bigcup_{j \in \mathbb{Z}} V_j = L^2(\mathbb{R}). \]

Condition I states that a multiresolution analysis comprises of a sequence of approximation spaces \( V_j \) where \( 0 < \cdots < V_{-1} \subset V_0 \subset V_1 \subset \cdots \subset L^2(\mathbb{R}) \).

Condition II shows that the approximation spaces are traversed by function \( \varphi \), which is known as the scaling function of the multiresolution analysis, so diverse selections for \( \varphi \) produces different multiresolution analysis and

\[ \| \varphi \|_2 = \left( \int_{-\infty}^{+\infty} |\varphi(x)|^2 \, dx \right)^{1/2} = 1. \]

For all \( j, k \in \mathbb{Z} \), the dilation, translation and normalization is denoted as

\[ \varphi_{j,k}(x) = 2^j \varphi(2^j x - k) \quad (3.32) \]

Condition II and III combined indicate that \( \{ \varphi_{j,k} : k \in \mathbb{Z} \} \) is an orthonormal basis for \( V_j \) for all \( j \in \mathbb{Z} \). For every \( j \in \mathbb{Z}, W_j \) is defined to be the orthogonal complement of \( V_j \) in \( V_{j+1} \). It entailsthat \( V_j \perp W_j \), \( V_j \ominus W_j = V_{j+1} \). Applying (3.29) recursively for \( J > J_0 \) complies that \( V_j = V_{j_0} \oplus W_{j_0} \oplus \ldots \ominus W_{j-1} \) where all the involved subspaces are orthogonal. Continuing the decomposition (3.26), and letting \( J_0 \to -\infty \) and \( J \to +\infty \) yields

\[ L^2(\mathbb{R}) = \bigoplus_{j \in \mathbb{Z}} W_j \quad (3.33) \]

It means that \( W_j \) is orthogonal. There exists a function \( \psi(x) \in W_0 \) such that \( \{\psi(2x - k)\}_{k \in \mathbb{Z}} \) is an orthonormal basis for \( W_0 \). So according to the multiresolution analysis attributes, the whole collection \( \{\psi_{j,k} : j, k \in \mathbb{Z}\} \),

\[ \psi_{j,k}(x) = 2^j \psi(2^j x - k) \quad (3.34) \]

is an orthonormal basis for \( L^2(\mathbb{R}) \). \( \psi(x) \) is called wavelet function.

An example of space \( V_j \) fulfilling the above-mentioned condition \( V_j = \{f \in L^2(\mathbb{R}) : f \) is constant on \([2^{-j}k, 2^{-j}(k + 1)]\), \( \forall j, k \in \mathbb{Z}\}, and here it is known as the
Haarmultiresolution analysis, and a likely choice for \( \varphi \) is the indicator function for \((0, 1)\). In this instance \( \varphi \) is known as the Haar scaling function of the multiresolution analysis and the function \( \varphi(x - k) \) has an identical graph as \( \varphi \), but transformed \( k \) units to the right. Since \( \varphi(x - k) \) is discontinuous at \( x = k \) and \( x = k + 1 \), and the scopes over a bounded set, so each element of \( V_0 \) is zero outside a bounded set. It entails that \( \varphi \) has a finite or compact support (the other Daubechies constructions of the wavelet system are compactly supported and continuous). Also for \( k \neq k' \), \( \varphi(x - k') \) and \( \varphi(x - k) \) have disjoint supports, therefore the set \( \{ \varphi(x - k) \}_{k \in \mathbb{Z}} \) is an orthonormal basis for \( V_0 \).

Basically, the Haar scaling function produces the subspaces \( V_j \), which are discontinuous in the set of integer multiplied by \( 2^{-j} \). The scaling condition for the Haar system (special case of Daubechies system) satisfies the condition III, so the Haar system of \( V_j \) satisfies all the properties of a multiresolution analysis.

With respect to the premises of multiresolution analysis, \( \{ \varphi(x - k) \}_{k \in \mathbb{Z}} \) is an orthonormal basis for \( V_0 \).

Checking the first and second conditions it means that \( \{ \varphi(2x - k) \}_{k \in \mathbb{Z}} \) is a basis of \( V_1 \), and substituting \( 2x = y \) it becomes

\[
\int_{-\infty}^{+\infty} \varphi (2x - k)\,dx = \frac{1}{2} \int_{-\infty}^{+\infty} \varphi (y - k)\,dy = \frac{1}{2} \int_{-\infty}^{+\infty} \varphi (\bar{x}) \,d\bar{x}, \tag{3.35}
\]

Then the scaling function has a mass of one, i.e. \( \int_{-\infty}^{+\infty} \varphi (x)\,dx = 1 \).

Equation (3.35) above gives \( \int_{-\infty}^{+\infty} \varphi (2x - k)\,dx = \frac{1}{2} \). Consequently, \( \| \varphi(2x - k) \|_{L^2} = \frac{1}{\sqrt{2}} \), and the normalized case is \( \varphi(2x - k) = \sqrt{2} \varphi(2x - k) \). Therefore, \( \{ \varphi(2x - k) \}_{k \in \mathbb{Z}} \) is an orthonormal basis of \( V_1 \). Algorithmically \( \{ 2^j \varphi(2^j x - k) \}_{k \in \mathbb{Z}} \) is orthonormal basis for \( V_j \).

Furthermore, for the wavelet space \( W_j \), \( \{ 2^j \varphi(2^j x - k) \}_{k \in \mathbb{Z}} \) is also an orthonormal basis.
The Scaling and wavelet functions is described below:

Subsequently, \( V_0 \subset V_1 \), any function \( f \) in \( V_0 \) can be expatiated in respect to the basis function for \( V_1 \). Therefore in specific the scaling function \( \varphi(x) \in V_0 \) and the wavelet function \( \psi \in V_1 \) can be expressed as:

\[
\varphi(x) = \sum_{k \in \mathbb{Z}} h_k \varphi_{1,k}(x) = 2^\frac{1}{2} \sum_{k \in \mathbb{Z}} h_k \varphi(2x - k) \quad (3.36)
\]

and

\[
\psi(x) = 2^\frac{1}{2} \sum_{k \in \mathbb{Z}} g_k \varphi(2x - k) \quad (3.37)
\]

where \( h_k = \int_{-\infty}^{+\infty} \varphi(x) \varphi_{1,k}(x) dx, g_k = \int_{-\infty}^{+\infty} \psi(x) \varphi_{1,k}(x) dx \) and \( \sum_{k \in \mathbb{Z}} |h_k|^2 = \sum_{k \in \mathbb{Z}} |g_k|^2 = 1 \).

Equation (3.36), (3.37) represents the scaling and wavelet equation, respectively. According to Daubechies, for a scaling function with compact support only a finite number of coefficients \( h_k \) and \( g_k \) are non-zero.

Hence, for the set of wavelets with compact support, scaling equation represents as:

\[
\varphi(x) = 2^\frac{1}{2} \sum_{k=0}^{D-1} h_k \varphi(2x - k) \quad (3.38)
\]

where \( D \in \mathbb{N} \) is known as the wavelet genus, and the numbers \( h_0, h_1, \ldots, h_{D-1} \) are called filter coefficients.

From Albeverio S and Skopina M [5], \( \psi \in W_0 \subset V_1 \), the same property holds for \( \psi \)

\[
\psi(x) = 2^\frac{1}{2} \sum_{k=0}^{D-1} g_k \varphi(2x - k) \quad (3.39)
\]

where \( g_k = (-1)^k h_{D-1,k} \) for \( k = 0, 1, \ldots, D-1 \), and \( \sum_{k=0}^{D-1} h_k g_k = 0 \).

Daubechies calculated this relationship and obtained discrete values for the filters \( h_k, g_k \) according to the attributes enforced to the multiresolution analysis. So the filters are unambiguously linked to the choice \( \varphi, \psi \) for the wavelet basis.

The inherent property of wavelets that allows for compression of data is explained here. The scaling function \( \varphi \) has the approximation property the polynomials up to order
$N - 1$ are reproduced exactly, $\sum_k M_k^p \varphi(x - k) = x^p$, for $p = 0, ..., N - 1$, where $N$ is the order of multiresolution analysis, $M_k^p$ is the $p$th moment of $\varphi(x - k)$. Take $x^p \in V_j$, since $V_j \perp W_j$, so $\langle x^p, \psi_{j,k} \rangle = 0$, for every wavelet function $\psi_{j,k}$ and

$$\int_{-\infty}^{+\infty} x^p \psi(x) dx = \sum_k M_k^p \int_{-\infty}^{+\infty} \varphi(x - k) \psi(x) dx = 0,$$

(3.40)

for $p = 0, ..., N - 1$. This explains that the wavelet $\psi$, associated to the scaling function $\phi$, has $N$ vanishing moments.

Based on the attributes posited by the Multiresolution framework, Scaling and wavelet functions with compact support are particularly suited for data decomposition and reconstruction in different levels of resolution.

### 3.4.2. Wavelet Feature Extraction

The wavelet transform provides a reliable platform suitable for image management because of its valuable features. The attributes that makes the wavelet transform valuable are; ability of compacting and transforming almost all the signal’s energy into few coefficients, a phenomenon known as energy compaction. The wavelet transform also has the ability of capturing and representing efficiently both the low frequency components and high frequency transients (like image edges). The adjustable resolution decomposition with nearly all uncorrelated coefficients and the ability to transmit progressively, hence, facilitating image reception at different qualities are also some of these properties.

Nearly all images match signals which have high-frequency constituent of short length (details) and low-frequency constituent of long length (approximations). Taking these into consideration, a multiresolution analysis operates best with this sort of signals. The 1-D wavelet should be the first step in the multiresolution analysis.

DWT has been extensively employed as a quick algorithm for finding the wavelet transform of signal tested in discrete time. The DWT essentially functions by examining the
signal through decomposition of the signal into rough approximation and detailed information, this is made possible by the use of high-pass and low-pass filters in the frequency domain.

Given signal $v(t) \in L^2(\mathbb{R})$, the DWT approximation coefficients and detailed coefficients are evaluated as:

$$
\sum_m h(m - 2k) cA_{j-1}(m)
$$

(3.41)

and

$$
cD_j(k) = \sum_m g(m - 2k) cA_{j-1}(m)
$$

(3.42)

where $j$ represents the level of decomposition, $k$ denotes time location, $m$ the number of samples, $h(*)$ and $g(*)$ are the half-band low-pass filters and high-pass filters, respectively. Observe that at each level $j$, the approximation coefficients alone are filtered allowing for the detail coefficients to be unaltered. For instance, take the DWT decomposition tree with three levels for signal $v(t)$ is shown in Figure 3.3.

By $j$ level decomposition the coefficients $cA_j$, $cD_j$, $cD_{j-1}$, ..., $cD_1$ can be assumed as the possible potential features to discriminate the different sort of defective signals. From the observation $cD_1$ can be rejected, as it mainly contains noise and no valuable information. This notion can be extended to multiresolution image analysis. For the purpose of image

![Figure 3.3. Three-level DWT decomposition tree](image-url)

71
processing, the 1-D DWT is stretched to 2-D DWT. The 1-D decomposition is process is performed along both row and column axes. This results in a total of four sub-band images HH (High-High), HL (High-Low), LH (Low-High) and LL (Low-Low). The different sub-band images hold different information and can be processed separately employing different algorithms. The sub-band image LL is equivalent to the lowest frequencies with the approximation co-efficients (cA). Here the background intensities and smooth information of the image is found. The sub-bands HL, LH, HH hold the image detail information i.e. the detail coefficients cV, cH and cD respectively. The sub-band HL contributes the vertical high frequencies (horizontal edges), LH contributes the horizontal high frequencies (vertical edges) and HH contributes the high frequencies in both directions (corners and diagonal edges).

Due to large dimensions of feature space, the SFFS method should be exploited to reduce the number of inputs.

3.4.2.1. Daubechies Family

In this segment the emphasis is on the particular instance of Daubechies wavelet transform constructed on the scaling and wavelet function with 2 vanishing moments which describes the filters having \( D = 4 \) coefficients represented as Db2. The scaling filters are expressed as

\[
\begin{align*}
    h_0 &= \frac{1 + \sqrt{3}}{4}, \quad h_1 = \frac{3 + \sqrt{3}}{4}, \quad h_2 = \frac{3 - \sqrt{3}}{4}, \quad h_3 = \frac{1 - \sqrt{3}}{4}.
\end{align*}
\]

As wavelet function is orthogonal to scaling function, the scaling and wavelet filter coefficient are linked as:

\[
\begin{align*}
    g_0 &= h_3, \quad g_1 = -h_2, \quad g_2 = h_1, \quad g_3 = -h_0
\end{align*}
\]

With this event, it can be deduced as:

\[
    h_0^2 + h_1^2 + h_3^2 = 1 \quad (3.43)
\]
\[ h_0 h_2 + h_1 h_3 = 0 \quad (3.44) \]
\[ h_0 + h_1 + h_2 + h_3 = 2 \quad (3.45) \]

where \( g_0^2 + g_1^2 + g_2^2 + g_3^2 = 1, \ g_0 + g_1 + g_2 + g_3 = 0. \)

Equation (3.43) and (3.44) matches the orthogonality of scaling functions. Equation (3.45) depicts the dilation property.

Using filters construction for Db2 scaling function, constructed on the calculation of its values on integers, and subsequently the dyadic grids, the scaling and wavelet function will be explained in this subsection. For the broad basis of construction founded on Fourier transform of the wavelet function and its attributes a detailed explanation is given by Ingrid Daubechies. Boggess A and Narcowich FJ [18] have shown how to calculate the values of \( \varphi \) at all dyadic points \( x = \frac{t}{2^n} \). This process is replicated in the following steps:

**Step 1. Calculation of \( \varphi \) at all the integer values**

For Db2, the scaling function is nonzero only on the interval \( 0 < x < 3 \). So \( \varphi(0) = \varphi(3) = 0 \). \( \varphi(1) \) and \( \varphi(2) \) are the only nonzero values at integer points. So for \( x = 1 \) and \( x = 2 \), the scaling equation

\[ \varphi(x) = \sum_k h_k \varphi(2x - k) \quad (3.46) \]

infers that \( \varphi(1) = h_0 \varphi(2) + h_1 \varphi(1) \), and \( \varphi(2) = h_2 \varphi(2) + h_3 \varphi(1) \)

therefore

\[ (h_1 - 1) \varphi(1) + h_0 \varphi(2) = 0 \quad (3.47) \]

Moreover, to organize the normalization \( \int \varphi = 1, \sum_l \varphi(l) = 1 \) is needed. Hence

\[ \varphi(1) + \varphi(2) = 1 \quad (3.48) \]

Consequently the result of equation (3.47) and (3.48) in the case of the Db2 are

\[ \varphi(1) = \frac{1 + \sqrt{3}}{2} \text{ and } \varphi(2) = \frac{1 - \sqrt{3}}{2} \]

**Step 2. Calculate \( \varphi \) at the half integers**
For Db2, \( \varphi(x) = 0 \) for \( x \leq 0 \) and \( x \geq 3 \). The only needed calculation is \( \varphi\left(\frac{l}{2}\right) \) for \( l = 1, 3, 5 \). For \( x = \frac{l}{2} \) using the scaling equation,

\[
\varphi\left(\frac{l}{2}\right) = \sum_k h_k \varphi(l - k)
\]

indicates that \( \varphi\left(\frac{1}{2}\right) = h_0 \varphi(1) = \frac{(1+\sqrt{3})^2}{8}, \varphi\left(\frac{3}{2}\right) = h_1 \varphi(2) + h_2 \varphi(1) = 0, \varphi\left(\frac{5}{2}\right) = h_2 \varphi(3) = \frac{(-1+\sqrt{3})^2}{8} \).

**Step 3. Perform Iteration**

Going on with this process, calculating the values of \( \varphi \) at \( \frac{l}{4} \) is interchangeable, and then only \( x = \frac{l}{4} \) is needed in the scaling equation. Broadly, calculating \( \varphi \) at the values \( \frac{l}{2^{n-1}} \), extends to calculating the value of \( \varphi \) at \( x = \frac{l}{2^n} \). It is then possible to calculate the wavelet function \( \psi \) using the scaling function \( \varphi \) as

\[
\psi(x) = \sum_{k \in \mathbb{Z}} (-1)^k h_{1-k} \varphi(2x - k)
\]

**3.4.2.2. Sub-band coding**

The broad view of sub-band decomposition to various scopes is free from ambiguity, especially in the separable event. The concept replete in sub-band coding and multiresolution analysis is similar to the principle behind continuous wavelet transform (CWT). In sub-band analysis digital filtering methods is used to achieve a time scale representation of signal. The CWT is an association between different scales and signals of a wavelet with the frequency used as a degree of resemblance. The CWT is calculated by varying the scale of the window used for analysis, changing the window consequently, employing signal multiplication, and integrating throughout the analysis. However, DWT uses the filters of altered limit frequencies to examine the signal at different scales. Signals
are passed through a series of high-pass-filters to examine the high frequencies, and passed through a series of low-pass-filters to study the low frequencies.

The filtering processes alters the resolution of the signal, the extent of the extent of detailed information in the signal, the up-sampling and down-sampling processes alters the scale. Signal sub-sampling represents the reduction in sampling rate, or removal of certain samples of the signal. For instance, the number of samples in a signal is reduced $n$ times by the factor $n$ of the sub-sampling process.

Up-sampling represents the increase in sampling rates of a signal by addition of new samples to the signal. For instance, the number of samples in a signal is increased $n$ times by the factor $n$ of the up-sampling.

Usually, the DWT coefficients are sampled from the CWT on a dyadic grid. Here the signal is a discrete time function. The signal filtering represents the convolution of the signal with impulse reaction of the filter. The convolution procedure in discrete time is set as follows:

$$x[n] * h[n] = \sum_{k=-\infty}^{\infty} x[k] \cdot h[n-k] \quad (3.51)$$

where the dyadic grid is denoted by $s_0 = 2$ and $t_0 = 1$, yielding $s = 2^l$ and $t = k \cdot 2^l$ as described in [71]. The function is noted by $x[n]$, and $n$ is an integer.

An important attribute of the DWT is the connection between the impulse reaction of the high-pass and low-pass filters. The high-pass and low-pass filter are interdependent, and are connected by $g[L - 1 - n] = (-1)^n \cdot h[n]$. Here $g[n]$ denotes the high-pass filter, $h[n]$ represents the low-pass filters, and $L$ is the length of the filter.

The two filtering sub-bands can be denoted as

$$y_{\text{high}}[k] = \sum_n x[n] \cdot g[-n + 2k] \quad (3.52)$$

$$y_{\text{low}}[k] = \sum_n x[n] \cdot h[-n + 2k] \quad (3.53)$$
The reconstruction is simple as the orthonormal bases are formed by the half-band filters. The reconstruction follows the reverse order of the operation stated above. It involves unsampling the signal at each level by two, passing through the production filter $g[n]$, and $h[n]$ respectively, and subsequently the added up. It is noteworthy to state that the analysis and production filters are similar, with the exception of the time reversal. Hence, the reconstruction can be represented as $x[n] = \sum_{k=\infty}^{\infty} (y_{\text{high}}[k] \cdot g[-n + 2k]) + (y_{\text{low}}[k] \cdot h[-n + 2k])$, this shows that the orthogonal basis of the dyadic nature of multiresolution approximation, in that it is closely linked to the theory of applying the DWT employing filter banks. Moreover, there is an algorithmic link between $c_j$ and $c_{j+1}$, also between $d_j$ and $d_{j+1}$

$$c_k^{j+1} = \sum_n c_k^j h[2k - n]$$

$$d_k^{j+1} = \sum_n c_k^j g[2k - n]$$

where $h[nn] = hhh[-n]a$ and $g[n] = g[-n]$

This operation above can be repeated for further decomposition and is known as the sub-band coding. At each level of operation, the filtering and sub-sampling gives rise to one-half the number of samples and one-half the high-pass and low-pass filters and they are dependent on each other, and linked by:

$$g[n] = (-1)^n h[1 - n]$$

However the filters that produce the perfect reconstruction, are the ones developed by Daubechies, known as the Daubechies wavelets. For reconstruction Filters, the importance of using the right filter is indicated. The choice of filters actually determines the possibility of prefect reconstruction. Moreover, the shape of the wavelet used in the analysis is also dependent on the choice of filter used. Reconstruction is performed at each level and sub-image while the size of the reconstructed image is kept at par with the original image.
3.4.2.3. Features from DWT

After applying image enhancement on the 100 UE images using the algorithm described earlier in previous section, the discrete wavelet transform (DWT) is applied. The resultant wavelet coefficient constitutes a set of features. The wavelet coefficient bands makes up the features of the signal at specific resolution contingent on position and scale of the wavelet. Wavelet coefficients are fundamentally separated into high frequency and low frequency coefficients. The high frequency coefficients are additional divided into vertical, horizontal, and diagonal coefficients. The approximation coefficients from the low frequency coefficients according to level of wavelet applied supplies a reduced resolution depiction of the data image. The Haar, Daubachies-2 to Daubachies-6 wavelets are applied in this study for evaluation purposes.

Tree-structure decomposition can be beneficial in providing useful information for texture pattern discrimination [23]. Most texture has their vital classification information in the middle bands. Image sub-band decomposition is based on the energy requirement. Using the LL, LH; and HL sub-bands a fixed decomposition tree can be achieved. The HH band is not usually left undecomposed because of its unstable features.

3.4.3. Segmentation Process

The wavelet transform basically function by decomposing the signal as a linear procedure for various scaling functions and can be achieved by filter convolution with altering signal and scaling.

In the case of the discrete signals, the DWT was achieved through the distinguishability of time, displacement and scale parameters. The wavelet is a linear compounding of the scaling and transformation of the scaling function that meets the
requirement for differential equation with two-scales, by employing equation (3.53) and (3.54):

\[
\phi(x) = \sqrt{2} \sum_k h(k) \phi(2x - k) \tag{3.57}
\]

\[
\psi(x) = \sqrt{2} \sum_k g(k) \phi(2x - k) \tag{3.58}
\]

where \( h \) represents the low-pass filter and \( g \) denotes the high-pass filter. The filters which are quadrature mirror, and fulfills \( g(k) = (-1)^k h(1 - k) \).

The wavelet transform application in image processing requires a two-dimensional (2-D) wavelet vector product of \( \psi(x) \) and \( \phi(x) \) which are specified using the following equation:

\[
\phi(x, y) = \phi(x) \phi(y) \tag{3.59}
\]

\[
\psi^H(x, y) = \psi(x) \phi(y) \tag{3.60}
\]

\[
\psi^V(x, y) = \phi(x) \psi(y) \tag{3.61}
\]

\[
\psi^D(x, y) = \psi(x) \psi(y) \tag{3.62}
\]

where \( \phi(x, y) \) is a 2-D scaling function, \( \psi^H \), \( \psi^V \) and \( \psi^D \) are three 2-D wavelets, and \( H, V \) and \( D \) denotes the horizontal, vertical and diagonal directions, correspondingly.

The texture feature extraction is dependent largely on the mother wavelet and the number of levels; this is basically because diverse mother wavelet shows different compact support in lengths and regularities. In this research, they were selected to extract features from breast UE images.

3.4.4. Summary

The 2-D wavelet decomposition of an image is executed by relating one dimensional DWT laterally on the rows of the image first, and then the outcomes are decomposed with respect to the columns. This operation results in four decomposed sub-band images LL,
LH, HL, and HH. DWT are used on discrete data sets and yield discrete outputs. This transformation of image vectors by DWT is similar to the Fast Fourier Transform (FFT) - the Fourier method done on a set of discrete quantities. The use of first level decomposition is proposed. This scheme considered the family of orthonormal wavelets with compressed support. These define an orthonormal foundation for the space of square integrable function. The simplest Daubechies is used in this transformation. Daubechies wavelet construct is based on the multi-resolution analysis framework. The proposed method used Daubechies wavelet transform with 1 to 6 coefficients in experimentation. The Db1 corresponds to the orthonormal system for square integrable function defined by Alfred Haar [50]. The Haar wavelet transform is a clear illustration of Daubechies with 2 scaling filter coefficients. The Db1 satisfies multi-resolution analysis attributes and is used for approximate function at different levels of resolution.

In the first stage the DWT is applied after the images had been enhanced using stick filters. The resulting coefficients after applying the wavelet transform constitute a set of features.

The wavelet transforms the enhanced image into four different sub-bands, with the resulting images distributed into overlapping square neighborhoods. It has been noted elsewhere that the attributes of wavelet transform mean that the substantial information of an image can be extracted by removing the sub-band of the wavelet decomposed image that holds the lowest frequencies [114]. The wavelet components reconstructed separately yield the diagonal, horizontal and vertical edges. Level one decomposition is applied to the samples using the Db1 to Db6 wavelets. The coefficients from each level of decomposition are then considered as features. Research also shows that malignancy in breast images corresponds to high frequency coefficients of the image spectrum. Here they were linked to wavelet coefficients in the top levels of resolution. A clear cut method to detect unhealthy
tissues is to decompose the breast image by wavelet transform, subdue the low frequency sub-band and recreate the input data solely for the high frequency related coefficients.

3.5. Modified Mean Local Binary Pattern (MM-LBP)

Texture analysis is the evaluation of the pixel position and gray level intensity of those pixels. LBP is a computationally efficient local texture descriptor. LBP operator proposed by Ojala et al. in 1996 works on 3 X 3 pixel block of an image. The detail of LBP is given in section 3.3. Either the labels obtained or their statistics are used for further image analysis. The high dimensionality of LBP histogram and its sensitivity to noise have raised concerns in many applications. LBP ignores gray level differences. This work proposes a novel extension of LBP, the Modified Mean -Local Binary Pattern (MM-LBP). Dimensionality reduction is proposed in MM-LBP that also tolerates noise better than the original LBP.

3.5.1. Proposed Methodology

A theoretically and computationally simple approach which is robust in term of gray scale variation is introduced in the proposed work. The proposed operator is an extension of LBP operator that represents the local texture information taking only the least details. LBP is based on the correlation between the center pixel $g_0$ and each of its neighboring pixels $g_1, g_2, g_3, g_4, g_5, g_6, g_7$ and $g_8$. As the neighborhood consists of 8 pixels, $2^8$ different labels are obtained depending on the gray value in the center pixel and its neighbors. The proposed MM-LBP considers neighbors $g_1, g_2, g_3, g_4$ in the four orientations 0°, 45°, 90°, 135° instead of eight neighbors for the pattern calculation, also the neighbors are correlated with the mean of the four neighbor pixels, taking the mean as the threshold. Figure 3.4 shows the neighbors considered for the pattern classification for LBP and the proposed method. LBP of a 3X3 image considers the neighbors in all the eight orientations, the MM-LBP aims at reducing the
number of labels considering only the four orientations 0°, 45°, 90°, 135°. In general, various neighborhood shapes such as circle, square and rectangular arrays are taken due to their ease of implementation. The neighbors are always centered at a center pixel. The center of the sub-image is moved from pixel to pixel in the process of MM-LBP calculation. The four neighbors from the top left corner are processed.

![Figure 3.4: Neighborhood taken (a) LBP (b) MM-LBP](image)

In MM-LBP, the pixel values are compared with the mean of the neighbors taken. \( g_p - g_c \) of LBP is invariant to the mean gray value but not to other changes in the gray levels. Sonoelastographic images require anisotropic information but do not require rotation invariance. As in the case of basic LBP only the signs of the differences are considered for texture classification:

\[
 t(s(g_1 - \mu), \ldots, s(g_4 - \mu)) \quad (3.63)
\]

where \( \bar{g} = \frac{\sum_{i=0}^{n} g_i}{n} \), \( n=4 \) for MM-LBP.

The thresholding function \( s(z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0. \end{cases} \)

The proposed operator is obtained by summing the differences obtained from eqn 6.1 weighted by powers of two.

The operator is defined as

\[
 \text{MM-LBP} \left( x_c, y_c \right) = \sum_{p=0}^{N-1} s(g_p - \mu)2^p \quad (3.64)
\]
where \( N = 4 \). Eqn 3.64 results in \( 2^N \) distinct values since the signs of differences of the neighborhood are interpreted as \( N \) bit binary number. Considering only four neighbors reduces the LBP labels significantly lower, than the basic LBP and the interpretation has been reduced from 8 bit to 4 bit binary number thus reducing the dimensionality. This is better suited as region descriptors. The occurrence frequency of the MM-LBP of an image represents the texture information of the image where the textures with similar labels are grouped together.

### 3.5.2. Global Descriptor

A global descriptor for the representation of the obtained pattern is the histogram. Histogram is the statistical occurrence of the similar patterns. Histogram can be formed from the whole image or from a sliding window around a center pixel, or from stationary windows or equally sized windows. Histogram of the MM-LBP and LBP codes are taken to measure the texture pattern in the image. Distance measures such as chi-square, histogram difference can be used to differentiate the textures as to which class the texture belongs. In sonoelastographic breast images there are two textures tumor and non-tumor that can be distinguished through the histogram.

### 3.5.3. Noise Sensitivity

Sonoelastography images have noise and artifacts introduced during the compression for obtaining the images. The noise present in the post compression images are additive Gaussian noise. LBP is sensitive to noise which is a critical issue for applications in medical field. To explain the changes in the LBP and the MM-LBP due to the introduction of noise a 3X3 image is taken as an example. Figure 3.5 shows the calculation of the LBP and the variation of the LBP code when the value of \( g_1 \) changes from 30 to 29. The proposed MM-LBP does not show any change in the local texture information obtained as in Figure 3.6.
The Figures 3.5 and 3.6 also show the dimensionality reduction. Though the two bit pattern looks similar, Figure 3.5 shows a small change in the input image results in a change
in the output LBP, while the same change in the input image of Figure 3.6 does not affect the
output from MM-LBP.

This proves that MM-LBP is insensitive to noise. MM-LBP is more robust to noise
when there are negligible changes in the pixel values of the input image.

3.5.4. Discriminability

Discriminability is the power of distinguishing one object from another. Due to the
discriminative power and computational simplicity, LBP is widely used in many areas of
research. MM-LBP aims at improving the discriminative power and robustness of the
operator. MM-LBP distinguishes tumor from non-tumor without any overlapping. The
discriminative power of the LBP and MM-LBP is explained in the Figures 3.7, 3.8.

In Figure 3.7 if the center pixel is 60 in place of 30 the binary code would be
01000000 and 11010011, so the LBP value would be 64, 211 respectively. The MM-LBP
does not depend on the center pixel so the MM-LBP values would be the same 5, 9. This

![Figure 3.7: Poor Discriminative power of LBP](image)

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>54</td>
<td>15</td>
<td>123</td>
</tr>
<tr>
<td>14</td>
<td>57</td>
<td>55</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>30</td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>29</td>
<td>123</td>
</tr>
<tr>
<td>14</td>
<td>190</td>
<td>155</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>255</td>
</tr>
</tbody>
</table>

Binary Code
11010011
LBP : 211

Binary Code
11010011
LBP : 211
shows that in few cases the LBP identifies the difference in texture. The discriminability of the LBP fails in few cases as of one shown in Figure 3.7. The changes in the input image do not affect the LBP labeling, whereas the change affects the MM-LBP label as in Figure 3.8. Thus the discriminative power is improved in the proposed MM-LBP.

3.5.5. Global Thresholding

The simplest way to segment the object from the image is thresholding. Thresholding isolates the relevant objects of interest. The threshold differentiates the similar and dissimilar textures which is helpful in the tumor and non-tumor detection. The output MM-LBP labels obtained for an image is thresholded for the segmentation of the sonoelastographic breast images. The efficiency of the proposed operator is evaluated by varying the threshold value (T). The overall minimum dissimilarity value is set as a threshold in segmenting the images.
From the experiments a constant threshold value $T$ is obtained and is applied for the MM-LBP and LBP labels.

### 3.5.6. Post Processing

Morphological operations are applied to the thresholded image to provide meaningful discontinuities for detecting the tumor. Thresholding introduces noise in the binary images that leads to imperfections in delineating the tumor from the non-tumor tissue. The Morphological operators are used in removing those imperfections that helps in detecting the tumor. The morphological operators used are explained in Section 3.3. After applying few of the morphological operators such as dilation, erosion, opening, closing, etc., the resultant images gives the area with the tumor.