# Chapter Two

# Theoretical Background

Women need to be well-versed in the anatomy and physiology of their breasts so that early signs of possible abnormalities can be clearly recognized. This introductory chapter aims to demonstrate the significance of breast cancer study and to provide some elementary knowledge on the breast structure, its related pathologies and an overview of the various methods that will be used in the implementation sector of this project.

## 2.1 Female Breast

A fully formed female breast is essentially composed of four structures, as indicative in Figure 2.1:

- Lobules/glands
- Milk ducts
- Fat
- Connective tissue

The lobules group together into larger units called lobes on an average of (15-20) in each breast, protruding from the nipple/areolar area. The lobules are responsible for producing milk for the infant from the nutrients and water taken from the bloodstream.

The lobes empty into the milk ducts which course through the breast towards the nipple/areolar area. They then converge into 6-10 larger ducts called collecting ducts which enter the base of the nipple and connect with the exterior of the breast.

There is predominance of glandular tissue in the upper outer portion of the breast which is responsible for the tenderness in this region. It is also the site of half of all breast cancers.

Surrounding the lobes is breast fat which is almost always soft. Interestingly, the difference in density between glandular breast tissue and breast fat is also the basis for mammographic imaging. The breast is also supported by semi-elastic bands of tissue called Cooper's ligaments. These ligaments (along with the skin) stretch over time when the gravity pulls the breast down.

Arteries carry oxygen rich blood from the heart to the chest wall and the breasts and veins take de-oxygenated blood back to the heart. The axillary artery extends from the armpit and supplies the outer half of the breast with blood while the internal mammary artery extends down from neck and supplies the inner portion of the breast.

The consistency of the breasts during the menstrual cycle is greatly varying due to peaking levels of estrogen and progesterone. Right after menstruation, hormone levels are at their lowest and the breast becomes softer and less tender. This is the recommended time to perform BSE and to have a mammogram [5, 6, 7].

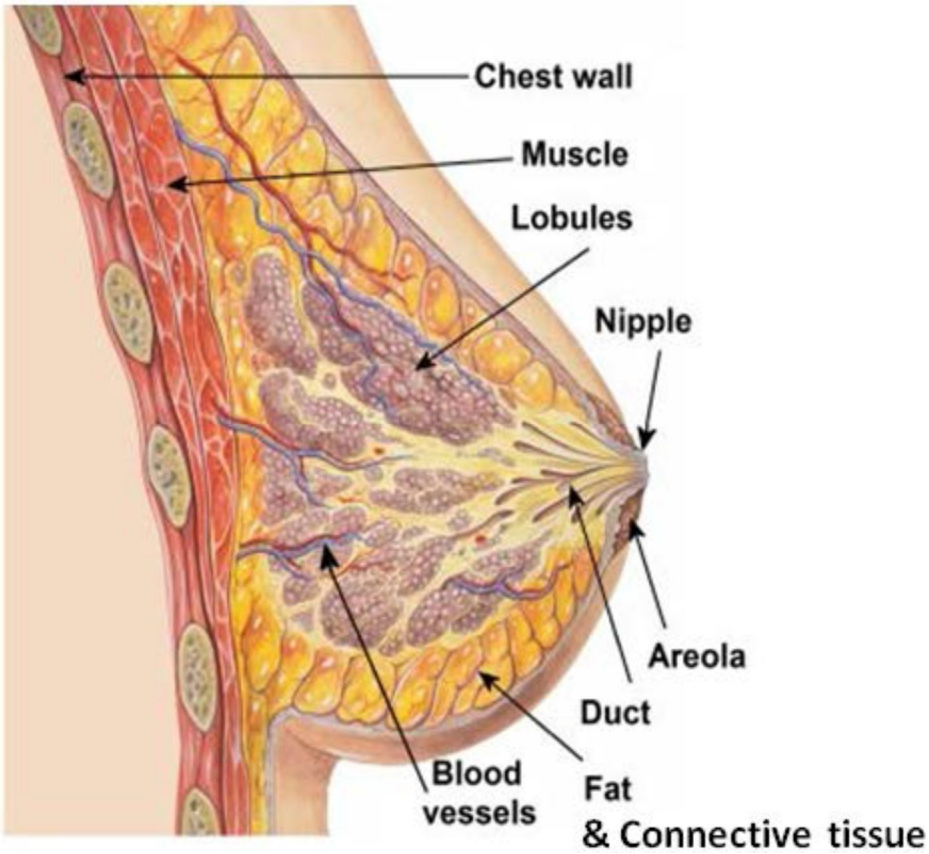



**Fig. 2.1 Anatomy of the Female Breast [8]**

### 2.1.1 Breast Cancer

The most notorious disease associated with the anatomy of the breast is breast cancer. It is considered a heterogeneous disease—differing by individual, age group, and even the kinds of cells within the tumors themselves. Breast cancer refers to life-threatening malignancies that develop in one or both breasts and is the most common form of cancer among women in developed countries. In fact, one in eight women will develop breast cancer [9, 10].

The human body is composed of cells that naturally process by dividing, growing and eventually dying. Most of the time cells divide and grow in an orderly manner. But sometimes cells grow out of control. This kind of cell growth forms a mass or lump called a tumor. Tumors are either benign or malignant [11]. Breast cancer does not have definite known causes. However, there are several common risk factors that may affect one's chances of developing breast cancer.

They include among others, early or late onset of menopause, family history of breast cancer, overweight/obesity, dense breast tissue and certain dietary factors. However, in over 75% of women with breast cancer none of these risk factors are present. In fact, the only clearly identifiable risks of developing breast cancer are gender and aging. Given this lack of predictability, preventative measures have been very difficult to determine [9, 10].

### 2.1.2 Breast cancer lesions

Given the definition of breast cancer, it is further classified by certain characteristics and types. Some lesions of breast cancer are as follows, calcifications, masses, and architectural distortions. Asymmetry between breasts can also be a breast cancer indicator.

Calcifications are one of the most important indicators of breast cancer. They are small size lesions that represent calcium deposits located in the breast tissue and are categorized into: macrocalcifications and microcalcifications. Normally, calcifications are not malignant or benign; but they tend to be associated with the malignant or benign tissue.

Macro-calcifications are coarse, scattered calcium deposits usually associated with benign conditions.

Microcalcifications (MCs) are tiny specks of calcium in the breast, usually found in an area of rapidly dividing cells with individual dimensions between 0.1-1 mm which makes it highly difficult to detect. A number of microcalcifications grouped together are termed as a cluster and are highly indicative of cancer. A cluster is defined as at least 3 microcalcifications within a 1-cm area [12].

Generally, larger, round and oval shaped calcifications with uniform size have higher probability of being benign, while smaller, irregular and branching calcifications, with heterogeneous size and morphology have higher probability of being malignant [9] [13]. Figure 2.2 shows several examples of microcalfications.

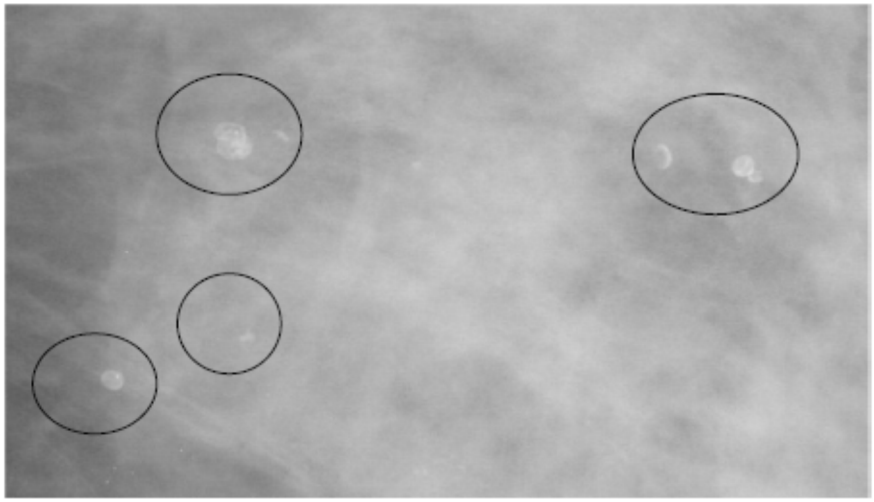


**Fig. 2.2: Microcalcifications (circled) [9]**

Architectural distortions refer to the deformation of the normal disposition of the parenchyma in a radiating or arbitrary pattern, without a visible centre or mass. They are very variable and, consequently, very difficult to detect [13]. Figure 2.3 shows an example of the mentioned type.

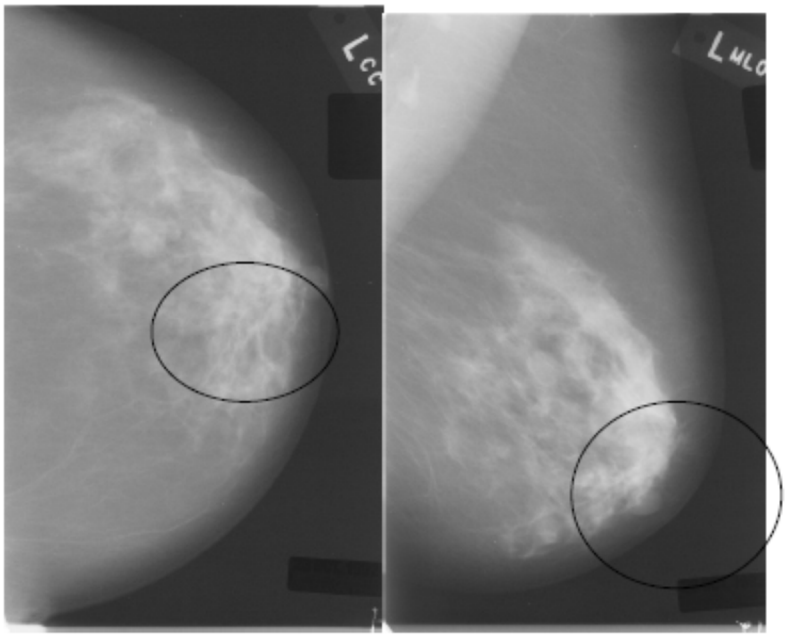


**Fig. 2.3: Side and overhead view of an architectural distortion (circled) [9]**

A final category, are masses. They are space-occupying lesions. They appear as dense regions of different sizes and properties and are categorized by their shape, density, and margins. They can be circular, oval, lobular or irregular/speculated and their margins can be as seen in Figure 2.4 and 2.5. The various types are as follows,

- Circumscribed, which are distinct well-defined boundaries;
- Obscured, which are hidden by superimposed or adjacent tissue;
- Micro-lobulated, which have undulating circular borders;
- Ill-defined, which are poorly defined scattered borders that taper into the background;
- Spiculated, which are radiating thin lines

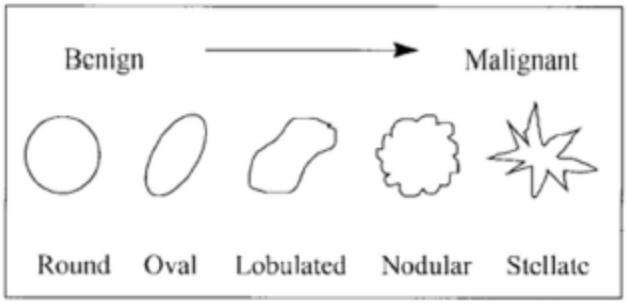



**Fig. 2.4: Morphologic spectrum of mammographic masses [13]**

Depending on the morphology, the masses have different malignant probability. The ill-defined and speculated borders have higher probability of malignancy whereas a benign process is usually associated with the presence of circular or oval masses. However, the great variability of the mass appearance is an obstacle to a correct mammography analysis.

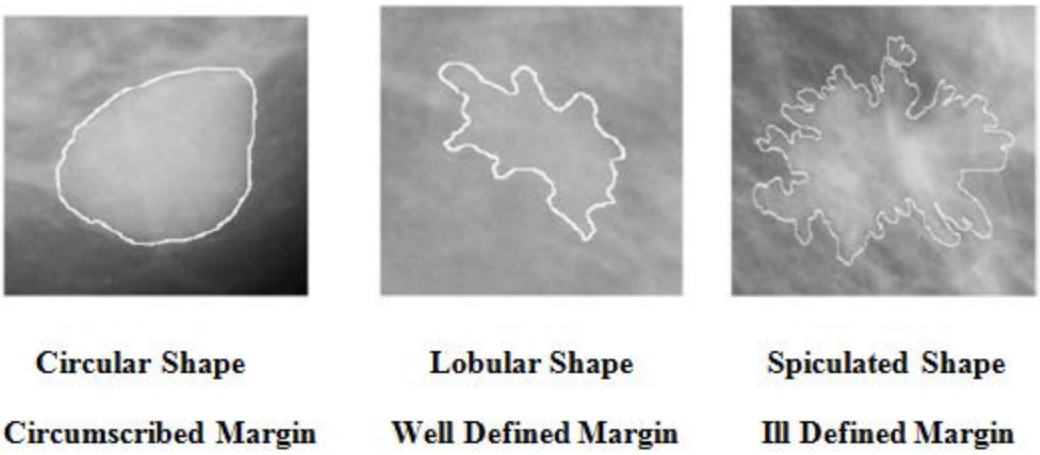



**Fig. 2.5: Mass examples with different shapes and borders [13]**

The densities of a mass include: high density masses that stand out sharply from the background tissue and low density masses that are faint and can be partially obscured. Equal density masses display similar intensities as other objects in the image whereas fat containing masses are faint due to low-density fat deposits in the mass [9].

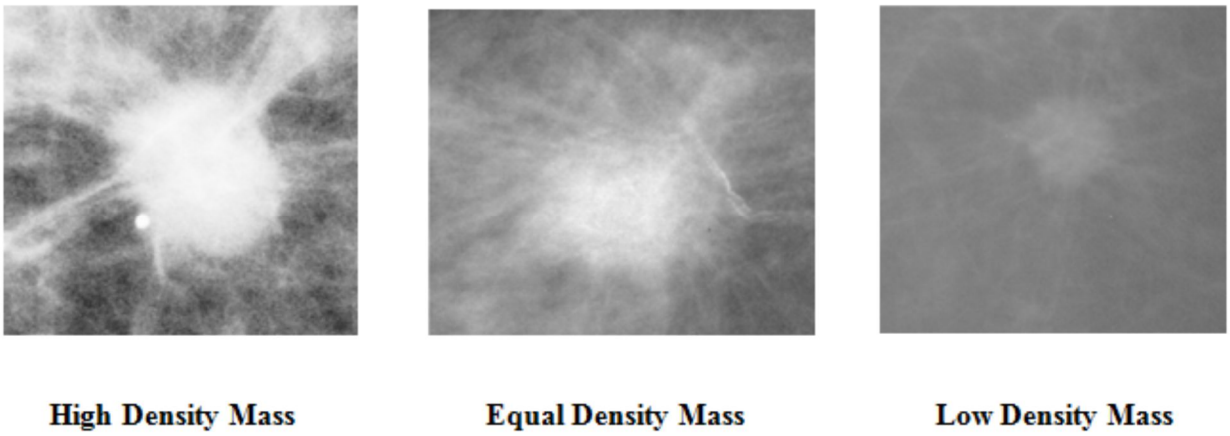



**Fig. 2.6: Mass examples with different densities [9]**

Masses are often indiscernible from the surrounding parenchyma because their features can be obscured or are similar to the normal inhomogeneous breast tissues. The gray levels of those inhomogeneous tissues in the breast could vary with the distribution of soft breast tissue. Thus, the difficulty in discrimination is escalated further by the fact that masses in digitized mammograms are similar to the glands, cysts or dense portion of the breast [14].

## 2.2 Mammography

Many consider early detection is the key to surviving breast cancer. The most effective method of screening is mammography. It is a diagnostic and screening tool that uses low-energy X-rays (Mo-K at around 30kVp) to examine the human breast. Its primary goal is to detect breast cancer at an early stage through previously mentioned cancer identifiers; characteristic masses, microcalcifications. Mammography results are normally expressed in terms of the BI-RADS (Breast Imaging-Reporting and Data System) Assessment Category. It has categories ranging from 0 (Incomplete) to 6 (Known biopsy – proven malignancy) [15]. Mammography is regarded the best selection for detecting breast cancer (benign/malignant) because it is able to show changes in the breast up to two years before a patient or physician can feel them [16].

## 2.3 Digital Mammography

Digital mammography is identical to conventional mammography but supersedes it by replacing x-ray film with solid-state detectors that convert x-ray into electrical signals. These electrical signals produce digital images of the breast that can be immediately stored or viewed on a computer screen [16]. It is regarded the better alternative because of its ability to manipulate the digital images for better visibility and contrast. Digital mammograms may take more views of each breast than film mammograms.

In addition, they use approximately 25% less radiation than film mammograms because smaller areas of the breast are imaged in each view. The radiation dose is even greatly reduced in women with larger and denser breasts [17] [18]. Regardless, digital mammography is roughly 1.5 to 4 times more expensive as conventional mammography and has to be developed to the detail resolution met within conventional mammography [19]. It only displays 3% of the information they detect, the main obstacles being the low contrast between the normal and abnormal glandular tissues especially in younger women.

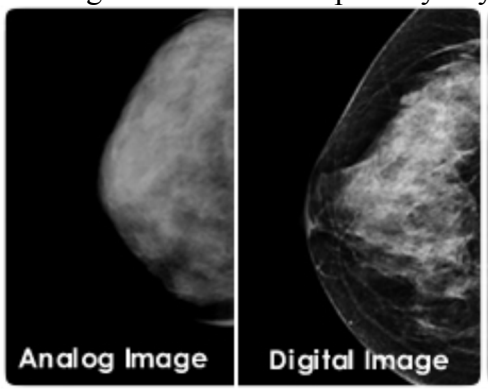



**Fig. 2.7: Comparing an analog with a digital image of the same breast [20]**

## 2.4 Mammographic Interpretation by Radiologists

The objective of mammogram interpretation is to detect all the abnormalities that are present on mammograms with no misdiagnoses of normal lesions. However, it is stated that10-30% of women with breast cancer, who undergo mammography, show negative mammograms in two-

thirds of which cancers are visible in retrospect [21].The judgments radiologists make are profoundly subjective on their training, experience and the criteria they work by [2]. The fact that visual search is not systematic but rather unique to each observer also contribute to the relatively high rates of variation in radiologic image interpretation.

The visual impair ability is further heightened by the fact that lesions are typically embedded in a background of anatomic noise, thus making detection, discrimination, and integration of distinct lesion features difficult.

A picture is worth a thousand words. Through its observation, it is regarded the most innate and convenient form for information interchange. Hence, vision is attributed the most advanced of the human senses which allow images to play an integral role in human perception.

Humans are limited to the extent of how much image information they can perceive. Unlike imaging systems that cover almost the entire Electromagnetic (EM) spectrum, ranging from gamma to radio waves, images are operatable from sources undecipherable by humans.

However, the analysis and interpretation of images by imaging systems are only an approximated (nearly degraded) version of the original. This is obviously due to unavoidable factors such as imaging geometry, lens aberration, wrong focus, unwanted motion and random sensor errors introduced into the image during the acquisition process [15] [16].

## 2.5 Digital Images

A continuous image, xij=f(i,j) which is discretized in both spatial coordinates and brightness is rendered a digital image. Therefore, in basic definition, a digital image is a matrix whose row and column indices identify a point in the image and whose corresponding matrix elemental value identifies the gray level at that point. Hence, for an M*N digital image, it is represented by the compact matrix form:

$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0,N-1) \\ f(1,0) & f(1,1) & \dots & f(1,N-1) \\ \vdots & \vdots & & \vdots \\ f(0,0) & f(0,1) & \dots & f(0,N-1) \end{bmatrix}$$

The elements of such a digital array are called image elements, picture elements or pixels. This is the definition of a grayscale image.

Grayscale images can be used to demonstrate variances in relative intensity for a given subject matter. Typically images will be defined by the range of values they contain. For example, an eight-bit grayscale image is one in which the pixel values range from 0 to 255. Likewise, a binary, or one-bit, image contains pixels which have values of zero or one. Figure 2.8 is an example of digitizing a continuous image.

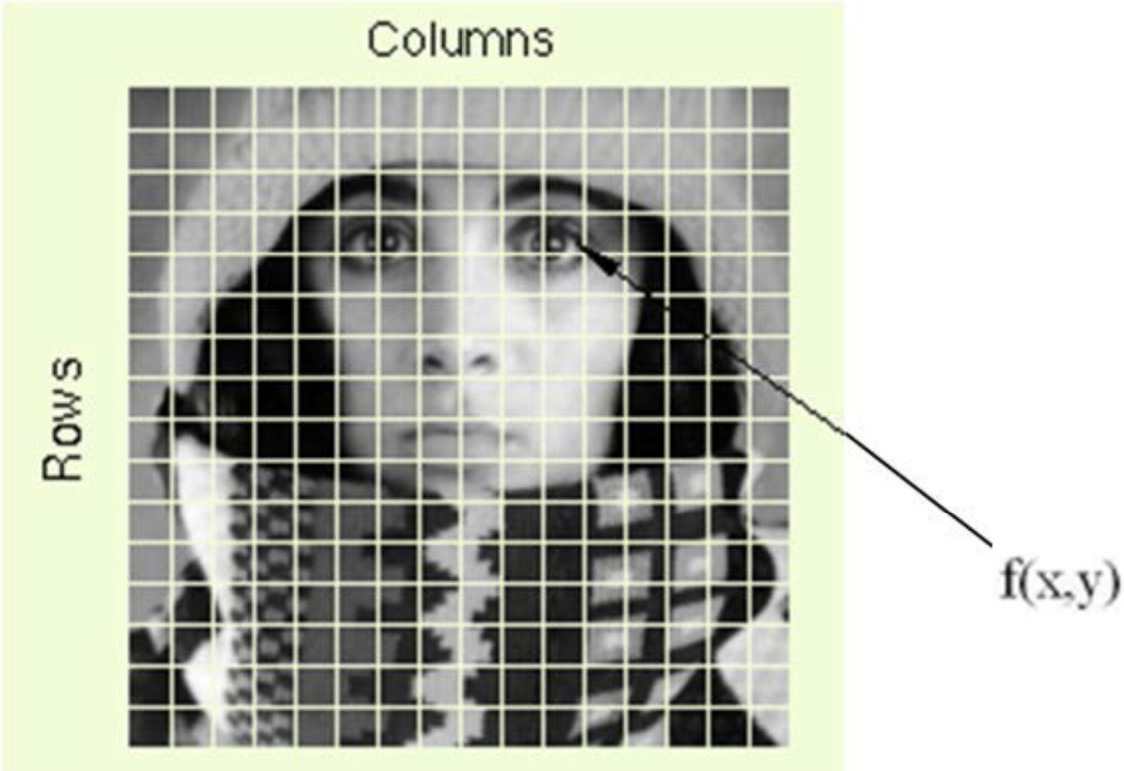



**Fig. 2.8: Digitization of a continuous image [22]**

Because the intensities captured on a mammogram x-ray film are records of the relative absorption of radiation, grayscale images are entirely suitable for digital mammogram images [9].

### 2.5.1 Digital Image Processing

As per the definition of a digital image in the preceding paragraph and considering the importance of images in the human world, the field of digital image processing culminates to define the processing of digital images by means of a digital computer. Two intrinsic application areas that nurture interest in digital image processing methods are:

- Improvement of pictorial information for human interpretation
- Processing of image data for storage, transmission, and representation for imaging system perception [23]

### 2.5.2 Applications of Digital Image Processing

Digital Image Processing serves to contribute towards many applications in the world today which are vast and innumerable. Regardless of the nature of the application, each utilizes the digital image processing techniques. Some of the few that are common are detailed below:

## X-ray Imaging

The use of x-rays dates back as one of the oldest sources of EM radiation for imaging. From the many applications it attributes to its best known use is the area of medical diagnostics. Be it for medical or industrial imaging, x-rays are generated using an X-ray tube, which is a vacuum tube with a cathode and anode. It is achieved by heating the cathode which releases free electrons. These electrons flow at high speed to a positively charged anode. When the electrons strike a nucleus, energy is released in the form of X-ray radiation. The energy (penetrating power) of the X-rays is controlled by a voltage applied across the anode, and the number of X-rays is controlled by a current applied to the filament in the cathode.

Computerized axial tomography (CAT) uses x-rays in medical Imaging. Each CAT image is a "slice" taken orthogonally through the patient. Numerous slices are thus, generated as the patient is moved in a longitudinal manner. The assembly of all the images constitutes a 3-D reconstruction of the patients' interior, with the longitudinal resolution being proportional to the number of slice images taken. Figure 2.9 demonstrates a typical CAT slice image.

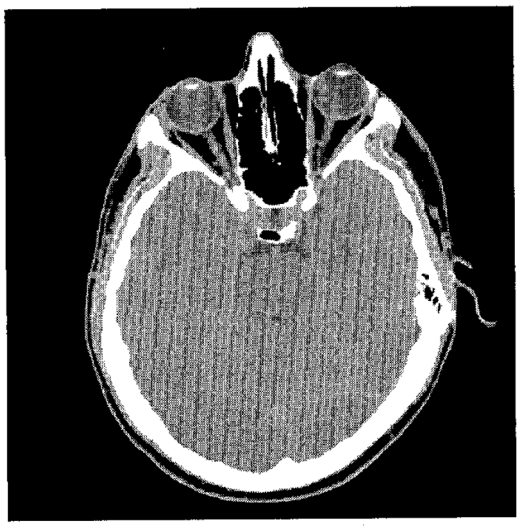


**Fig. 2.9: CAT slice image [23]**

## Gamma-Ray Imaging

Applications of imaging based on gamma rays among many others, include nuclear medicine and astronomical observations. In the medical field of nuclear medicine, a radioactive isotope that emits gamma rays as it decays is injected into the patient. The images are thus produced from the emissions collected by gamma ray detectors. Such images are used to locate sites of bone pathology, such as infections or tumors. An example of a Gamma-ray imaging is illustrated in Figure 2.10.

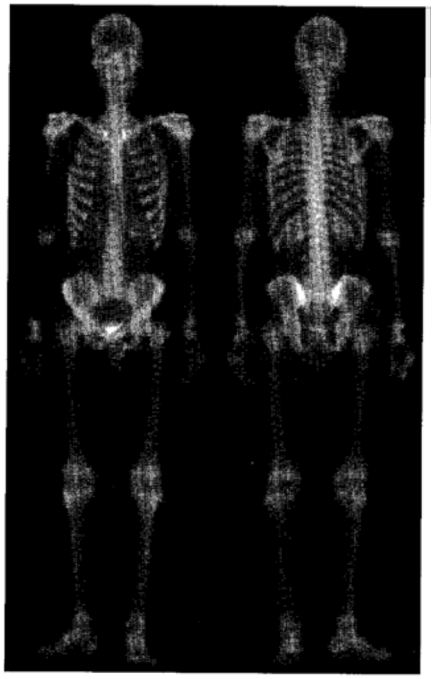


**Fig. 2.10: Example of Gamma-ray imaging for a bone scans [23]**

**"Sound" image applications**

Imaging using "sound" finds applications in geological exploration, industry, and medicine. For instance, geological applications use sound in the low end of the sound spectrum (hundreds of Hertz) while imaging in other areas use ultrasound (millions of Hertz) [23].

### 2.5.3 MATLAB and the Image Processing Toolbox

The name MATLAB stands for *matrix laboratory*. MATLAB was written originally to provide easy access to matrix software developed by the LINPACK (Linear System Package) and EISPACK (Eigen System Package) projects. Today, MATLAB engines incorporate the LAPACK (Linear Algebra Package) and BLAS (Basic Linear Algebra Subprograms) libraries, constituting the state of the art in software for matrix computation.

MATLAB is a high-performance language for technical computing. Integrating computation, visualization, and programming in a user-friendly environment, problems and solutions are easily expressed in familiar mathematical notation. Typical uses include the following:

- Math and computation
- Algorithm development
- Data acquisition
- Modeling, simulation, and prototyping

- Data analysis, exploration, and visualization
- Scientific and engineering graphics
- Application development, including graphical user interface building

It is an interactive system whose basic data element is an array that does not require dimensioning. This allows formulating solutions to many technical computing problems, especially those involving matrix representations, in a fraction of the time it would take to write a program in a scalar non-interactive language such as C or FORTRAN.

It features a family of add-on application-specific solutions called toolboxes. Very important to most users of MATLAB, toolboxes allow you to learn and apply specialized technology. Toolboxes are comprehensive collections of MATLAB functions (M-files) that extend the MATLAB environment to solve particular classes of problems.

The Image Processing Toolbox is a collection of functions that extend the capability of the MATLAB numeric computing environment. The toolbox supports a wide range of image processing operations, including:

- Neighborhood and block operations
- Filter design
- Image analysis and enhancement
- ROI operations [24]

## 2.6 Image Enhancement Techniques

Image enhancement techniques are of integral importance, as they improve the quality of an image, eliminate noise or error, or accentuate certain features The underlying principle of image enhancement is to enlarge the intensity difference between objects and background and to produce reliable representations of breast tissue structures. These operations can be quite useful in the development of digital mammogram analysis techniques.

**Histogram**

Histogram is a good indicator of image contrast. Its application in image processing techniques such as stretching, compression and segmentation is vast. It is a plot of the number of pixels in the image having a certain gray level value versus the range of gray level values present, irrespective of the image location. The horizontal axis displays pixel variations from the darkest on the left with the smallest value of 0 to the lightest on the right with the largest value of 255.Since histograms enable the task of stretching and compression, low contrast and high contrast can be manipulated by changing the value distribution of the image's histogram. Figure 2.11 is an example of a mammogram image with its corresponding histogram. The histogram shows the probability distribution for each gray level in the image. Since Figure 2.11 is an eight-bit image, the histogram shows 256 values.

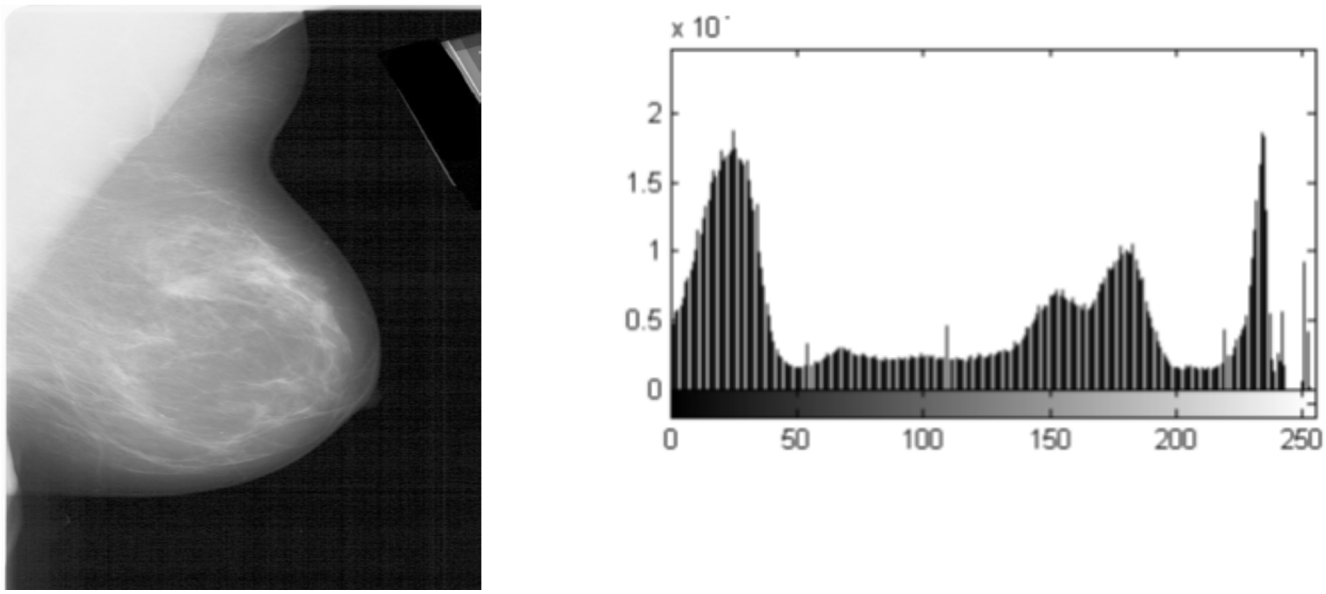



**Fig. 2.11: Mammogram image representing side view of left breast with corresponding histogram [9]**

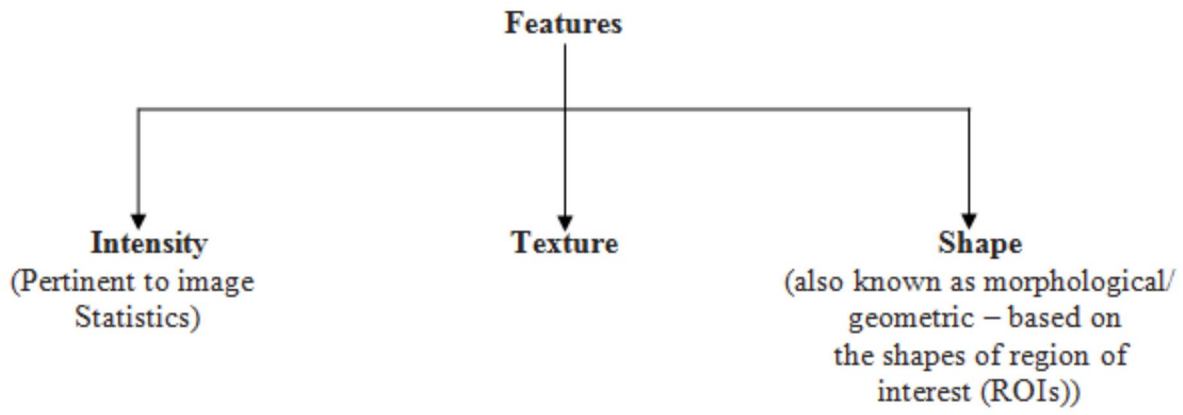
**Contrast Limited Adaptive Histogram Equalization**

If the region in the image has a relatively small intensity range, there is a possibility of image over-enhancement such as noise, since it is enhancement in itself. This can lead to some kind of undesired artifacts to appear on those regions. The amount of contrast enhancement for some intensity is directly proportional to the slope of the CDF (cumulative distribution function) function at that intensity level. Therefore, in order to limit this unwanted effect, contrast enhancement can be limited by limiting the slope of the CDF. The slope of CDF at a bin location is determined by the height of the histogram for that bin. Therefore if we limit the height of the histogram to a certain level we can limit the slope of the CDF and hence the amount of contrast enhancement [25].

The CLAHE algorithm is a widely used technique which results in contrast enhancement of medical images. The histogram is cut at some threshold and then equalization is applied. It is adaptive contrast histogram equalization method, where the contrast of an image is enhanced by applying CLAHE on small data regions called tiles rather than the entire image. The resulting neighboring tiles are then stitched back seamlessly using bilinear interpolation. The contrast in the homogeneous region can be limited so that noise amplification can be avoided [26].

## 2.7 Features

Features are simply defined as characteristics of the objects of interest in the image which constitutes informational values that characterize the image. There are innumerable varying features that can be extracted from a digital mammogram. As Cheng et al. [2] mentioned, features are classified mainly as follows:




When careful selection of features is made, it is sufficient enough to be representative of the maximum relevant information that the image has to offer for a complete characterization of the lesion [27]. In this proposed project, texture features will be used.

### 2.7.1 Texture features

According to Sklansky (1978), ‘‘an image region has a constant texture if a set of local properties in that region is constant, slowly varying or approximately periodic’’. The local statistics or property that is repeated over the textured region is called a texture element or texel [28].

Texture analysis of features of mammographic breast tissue is an issue of major importance in mass characterization. However, due to its inherent difficulty and fuzziness, it has not been investigated in depth with only a few studies available to support its exploration, in contrast to other mammographic diagnostic approaches [29].

#### 2.7.1.1 Gray Level Co-occurrence Matrices (GLCM)

Some of the textural features are contained into a class, based on spatial gray level dependence (SGLD) matrices, or gray-level co-occurrence matrices (GLCM). Co-occurrence matrices are two-dimensional histograms of the number of times that pairs of intensity values occur in a given spatial relationship. Therefore the GLCM is an advanced method of using different combinations of pixel brightness values which are the grey levels in the image.

Since a co-occurrence matrix contains a count of the number of times that a given feature (e.g. a given grey level) occurs in a particular spatial relation to another given feature, Figure 2.14 is an illustration of this concept.

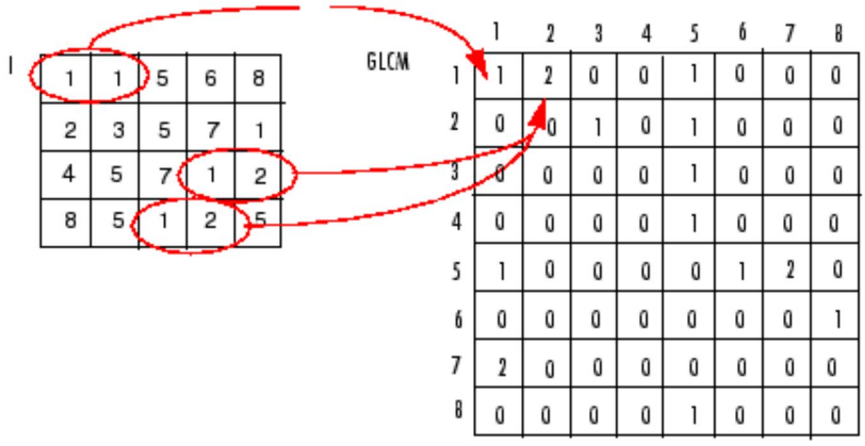



**Fig. 2.12: Example of how to generate a Co-occurrence matrix with eight grey levels**

For the calculation of values in the GLCM of the 4x5 image (I), element (1,1) in the GLCM contains the value 1 because there is only one instance in the image where two, horizontally adjacent pixels have the values 1 and 1. Element (1, 2) in the GLCM contains the value 2 because there are two instances in the image where two, horizontally adjacent pixels have the values 1 and 2. The other elements are computed this way to fill all the values in the GLCM.

In this project, the GLCM that will be calculated for all sample images will be calculated for each pixel in question and the neighboring pixel separated by distance 2 and along (0°) direction of the image. The chosen direction is one-dimensional picked for easier computational means.

The GLCM is constructed by observing pairs of image cells distance *d* from each other and incrementing the matrix position corresponding to the grey level of both cells. Therefore, it contains the frequency of a certain pair of pixels repetition in an image. An element of the GLCM matrix $P(i, j, d, \theta\theta)$ is hence, defined as the joint probability that the gray levels *i* and *j* occur separated by distance *d* and along $\theta$ direction of the image.

The element *(i , j)* of the square matrix is generated by counting the number of times a pixel with value *i* is neighboring to a pixel with value j and then dividing the entire matrix by the total number of such comparisons made. This derives four matrices for each given distance in the following directions (0°, 45°, 90° and 135°) as shown in Figure 2.15 [2] [30].

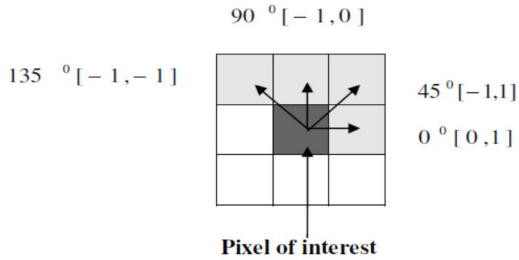



**Fig. 2.13: Adjacency of pixel in four directions (horizontal, vertical, left and right diagonals) [2] [30]**

## 2.8 ANFIS Classification

ANFIS is one of hybrid neuro-fuzzy inference expert systems and it works in Takagi-Sugeno-type fuzzy inference system, which was developed by Jang [31]. ANFIS has a similar structure to a multilayer feed forward neural network but the links in an ANFIS only indicate the flow direction of signals between nodes and no weights are associated with the links [32].

### 2.8.1 ANFIS Architecture

In ANFIS, Takagi-Sugeno type fuzzy inference system is used. The output of each rule can be a linear combination of input variables plus a constant term or can be only a constant term. The final output is the weighted average of each rule's output. Basic ANFIS architecture that has two inputs x and y and one output f is shown in Figure 2.16. The rule base contains two Takagi-Sugeno if-then rules as follows [33]:

$$Rule1 = If x\ is A_1\ and\ y\ is\ B_1\ , then\ f_1 = p_1 x + q_1 y + r_1 \qquad (2.1)$$

$$Rule2 = If x\ is A_2\ and\ y\ is\ B_2\ , then\ f_2 = p_2 x + q_2 y + r_2 \qquad (2.2)$$

Where $p_1, q_1, r_1, p_2, q_2, and\ r_2$ are the linear consequent parameters.

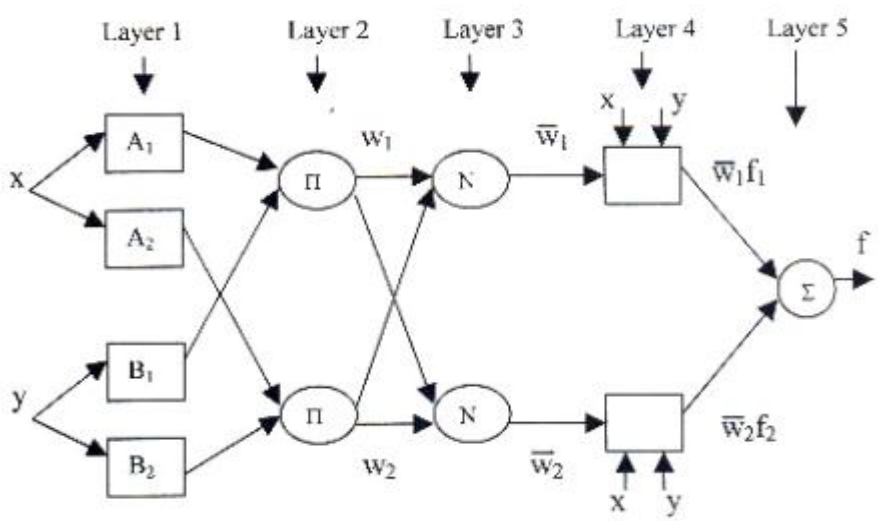



**Fig. 2.14: Basic structure of ANFIS [33]**

The node functions in the same layer are the same as described below:

**Layer 1:**

Every node, $\boldsymbol{i}$ in this layer is a square node with a node function as:

$$o_{1,i} = \mu_{A_i}(x), for\ i = 1,2 \qquad (2.3)$$

$$o_{1,i} = \mu_{B_{i-2}}(y), for\ i = 1,2 \qquad (2.4)$$

Where x is the input to node $\boldsymbol{i}$, and $A_i$(or $B_{i-2}$ ) is a linguistic label (such as "small" or "large") associated with this node. In other words, $o_{1,i}$ is the membership grade of a fuzzy set $\boldsymbol{A}$ and it specifies the degree to which the given input x satisfies the quantifier $\boldsymbol{A}$. The membership function for A can be any appropriate membership function, such as the Triangular or Gaussian. When the parameters of membership function changes, chosen membership function varies accordingly, thus exhibiting various forms of membership functions for a fuzzy set $\boldsymbol{A}$. Parameters in this layer are referred to as "premise parameters"[33].

**Layer 2:**

Every node in this layer is a fixed node labeled as Π, whose output is the product of all incoming signals:

$$o_{2,i} = w_i = \mu_{A_i}(x)\mu_{B_i}(y), for\ i = 1,2 \qquad (2.5)$$

Each node output represents the firing strength of a fuzzy rule [34].

**Layer 3:**

Every node in this layer is a fixed node labeled N. The $i_{th}$ node calculates the ratio of the rule's firing strength to the sum of all rules' firing strengths:

$$o_{3,i} = w_i = w_i/(w_1 + w_2), i = 1,2 \qquad (2.6)$$

Outputs of this layer are called "normalized firing strengths" [34].

**Layer 4:**

Every node i in this layer is an adaptive node with a node function as:

$$o_{4,i} = \overline{w_i} f_i = \overline{w_i}\ (p_i x + q_i y + r_i) \qquad (2.7)$$

Where $\overline{w_i}$ is a normalized firing strength from layer 3 and $\{p_i, q_i, r_i\}$ is the parameter set of this node. Parameters in this layer are referred to as "consequent parameters" [33].

**Layer 5:**

The single node in this layer is a fixed node labeled Σ that computes the overall output as the summation of all incoming signals:

$$overall\ output = o_{5,i} = \sum_i \overline{w_i} f_i = \frac{\sum_i w_i f}{\sum_i w_i} \qquad (2.8)$$

Thus an adaptive network, which is functionally equivalent to the Takagi-Sugeno type fuzzy inference system, has been constructed [34].

Other example of ANFIS with nine rules can be shown in Figure 2.8. Three membership functions are associated with each input, so the input space partitioned into nine fuzzy subspaces.

The premise part of a rule describes a fuzzy subspace, while the consequent part specifies the output within this fuzzy subspace [33].

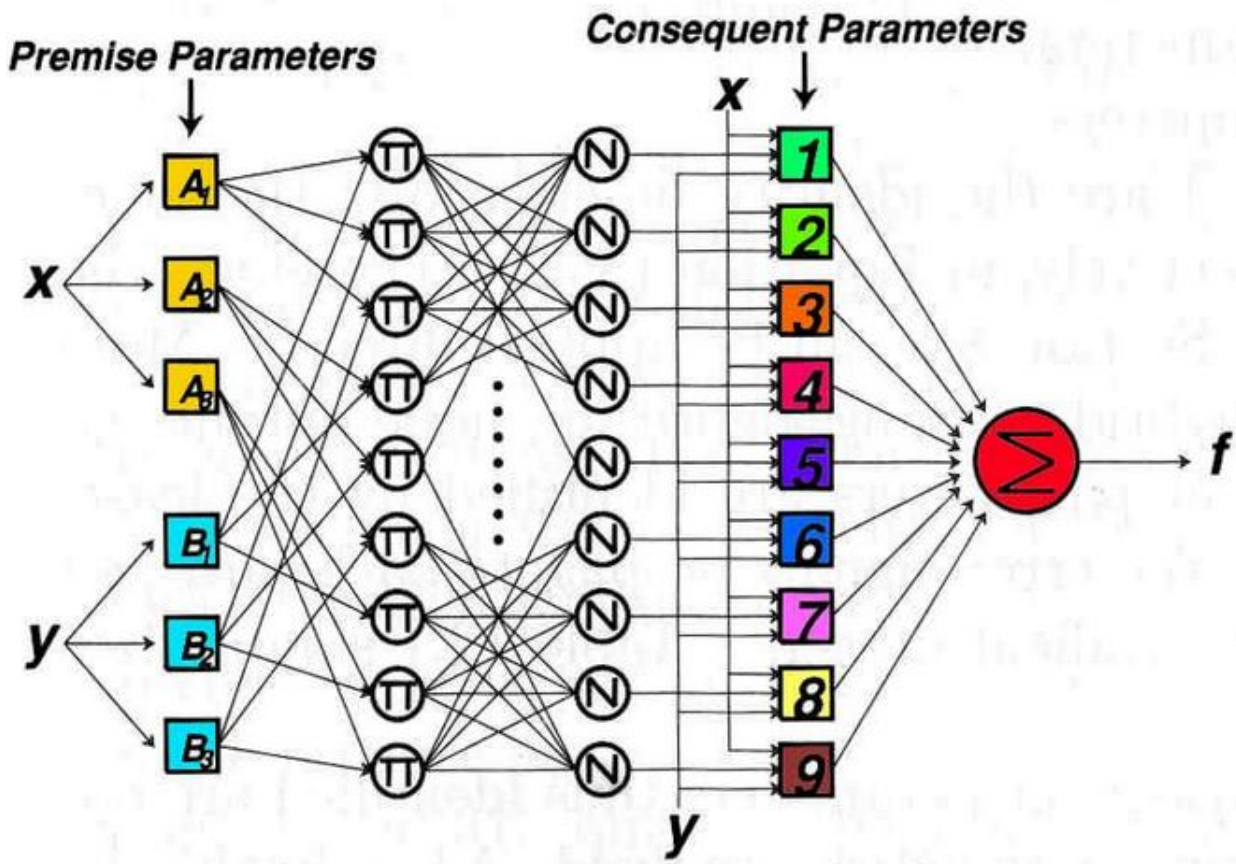



**Fig. 2.15: ANFIS Architecture with nine rules [34]**

### 2.8.2 ANFIS Learning Algorithm

From the proposed ANFIS architecture above (Figure 2.16), the output $\boldsymbol{f}$ can be defined as:

$$f = \frac{w_1}{w_1+w_2} f_1 + \frac{w_1}{w_1+w_2} f_2 \qquad (2.9)$$

$$f = \overline{w_1}(p_1(x) + q_1 y + r_1) + \overline{w_2}(p_2(x) + q_2 y + r_2) \qquad (2.10)$$

$$f = (\overline{w_1}\, x)p_1 + (\overline{w_1}\, y)q_1 + (\overline{w_1})r_1 + (\overline{w_2}\, x)p_2 + (\overline{w_2}\, y)q_2 + (\overline{w_2})r_2 \qquad (2.11)$$

The methods for updating the parameters are listed as below [33]:

i. Gradient decent only: All parameters are updated by gradient decent backpropagation.

ii. Gradient decent and one pass of Least Square Estimates (LSE): The LSE is applied only once at the very beginning to get the initial values of the consequent parameters and then the gradient descent takes over to update all parameters.

iii. Gradient and LSE: This is the hybrid learning rule [33].

Since the hybrid learning approach converges much faster by reducing search space dimensions than the original backpropagation method, it is more desirable. In the forward pass of the hybrid learning, node outputs go forward until layer 4 and the consequent parameters are identified with the least square method. In the backward pass, the error rates propagate backward and the premise parameters are updated by gradient descent [34].