Chapter II
LITERATURE REVIEW

2.1. Introduction

Image segmentation methods can be categorized into Edge (contour) based techniques and region-based techniques. Edge-based technique sets out with edge detection, and then graduates to a link up process that taps into curvilinear continuity, that finds the object boundaries and then locate the object itself by filling them in. A region based approach takes the opposite approach. For example, starting in the middle of an object and then “growing” outward until it meets the object boundaries. Region based technique tries to set image pixels to consistent image properties like texture, color and brightness.

The edge and region based approaches are not actually different from each other. The edges of regions can be delineated to be the contours. If the edge-based model is implemented then the regions can be specified from a contour-based technique. Region-based approaches readily seek to define a global objective function - such as Markov random fields or variant formulations. The benefit of having a global objective function is that conclusions are made only when information from the whole image is taken into consideration simultaneously. In the edge and region based techniques the distinguishing feature is of more significance, and the clustering factor or grouping factor is coded more normally in a given model[36, 106].

In most contour-based techniques, the first step often is to detect the edge locally. Afterwards, care is taken to improve the results by a global linking process that exploits curvilinear continuity, such as stochastic completion, relaxation techniques, or saliency networking. A major drawback of this technique is that the edge/no edge conclusions are made in an untimely manner. A very low threshold has to be set to be able to detect an
extended contour with very low contrast. This invariably leads to random edge segments being found everywhere in the image, thereby making the task of curvilinear link processing more difficult than if the raw contrast information was used. However, when dealing with images which have both textured and untextured regions as UE images do, boundaries must be detected using both contour and texture analysis. This is not so in the literature reviewed, and the concentration is always on one technique [28,90,120].

Certainly edge-detection may be sufficient for images without texture, but in a textured region it is a directionless convoluted network of contours. The dangers of using edge-detection have been noted before [16], but the additional problem of contours creating problems for texture analysis has not been discerned. Distinction techniques are based on measuring texture descriptors over local windows, and then measuring computational differences between window descriptors at different locations.

By and large, various image processing problems are designed to locate a group of pixels in an image that in some way fit together. Discovering such clusters is critical if measurements on objects represented in an image is the goal. In a medical image intended to measure the mass of the breast, there is a need to determine which pixels make up this mass. Finding these clusters of pixels is also an important step in various automated object detection tasks. There are several existing methods for selecting a proper threshold value for a segmentation problem.

The commonest method used is to set the threshold value synergistically. The user operates the value and revises the thresholding end result until the desired segmentation is achieved. The histogram is a useful instrument in setting up a suitable threshold value. If the idea is to establish more segments, the method defined can be stretched to apply compound thresholds.
The variance ultimately will be a function made up of more than a single threshold; hence multi-dimensional optimization is required to discover optimal thresholds. However, this is particularly clumsy, with computational complexity when the number of segments required is large. A supplementary real-world algorithm that reduces the variance within segments, frequently utilized is an iterative or step-function algorithm known as K-means clustering [72]. The goal of the K-means clustering algorithm is to split an image into K segments, decreasing the overall within-segment variance. The variable K is set prior to operating the algorithm.

2.2. Segmentation of Breast Tumor

The gold standard for breast imaging and tumor detection is mammography. Nevertheless, this imaging modality has some limitation like reduced sensitivity and specificity in dense breast, hence, other imaging modalities such as US and UE are frequently recommended to gain additional information missed with mammography.

US imaging has continued to gain more grounds in the area of breast cancer detection. Studies have revealed that US images can differentiate benign and malignant masses with a high accuracy. Use of this imaging modality can increase overall cancer detection by 17% as well as reduce unneeded biopsies by 40% [106]. US imaging modality are superior to mammography based on the fact that: it has no radiation, hence, it is safer for the patient and the radiologist in day-to-day clinical use [41]. It is also relatively cheaper and much faster than mammography and cost-effective for the low and middle income people.

Secondly, UE and US has better sensitivity and specificity than mammography especially in dense breast, and proves invaluable for women less than 35 years. High rate of false positive finding in mammography [13,133] is greatly reduced in US imaging were accuracy rate of simple cysts is about 96-100% [111].
Conversely, the mammography is less operator dependent than the sonography and sonoelastography, the sonogram and elastogram requires well trained radiologists with many years of experience to be interpreted accurately. Even among expert radiologist the issue of observer variations arises, hence, an automatic detection system or computer-aided diagnosis (CAD) is needed to help radiologists in their diagnosis [22]. Several CAD systems have been developed in order to help radiologists in diagnostic accuracy. Researchers have proposed several CAD approaches like the artificial neural network (ANN), linear discriminant analysis (LDA) and the support vector machine (SVM) [7, 27, 58, 121] for mass detection and classification of breast cancer.

In general terms the ultrasound CAD system comprises of four phases (i) preprocessing or image enhancement (ii) Region Of Interest (ROI) segmentation (iii) feature extraction and selection, and lastly (iv) classification. The image processing and segmentation components are left out in some CAD system. In these models, the texture feature are harvested right away from the image or ROI and used as inputs of the classifiers system [67]. Such kind of CAD system is simple, with low computational complexities. However, features extracted from the ROIs directly may not be able to support robust and precise execution.

2.3. Image Enhancement

The subject of image enhancement has been an area of rigorous research in medical imaging and enjoys broad attention. Image enhancement is simply refinement of an image to make its information more construable for the human viewer, and supply more robust input data for unsupervised image processing schemes. The chief aim of image enhancement is to alter the features of an image to produce a better image and make it more suited for a particular task. Techniques used in medical image enhancement renders diverse choices for
enhancing the quality of images. The imaging modality and the particular task greatly influences the choice of procedure used. Increasing the contrast of an image, followed by filtering to eliminate noise is a typical example of image enhancement. Image enhancement algorithms improve the quality of images by removing noise, blurring, and increasing the sharpness and contrast of medical images. There are several image enhancement techniques such as noise smoothing, Histogram equalization, Range compression, Contrast stretching, Filtering and enhancement in the frequency domain. The review of available image enhancement techniques shows that the various techniques can be classified into two broad groups: Spatial domain and Frequency domain.

Spatial based domain simply refers to the image plane, and methods in this grouping are based on the direct operations on the pixels in an image. The conceptual simplicity and lack of computational complexities gives the spatial based approaches an edge in real time image processing. However, in terms of cogency and unperceivability necessities the technique is found wanting. Frequency based domain refers to mathematical functions analysis with regard to frequency and are based on the direct modification of the transform coefficients of the image, like the Fourier transform and Discrete Wavelet Transform (DWT). The underlying principle is to “better” the image by modifying the transform coefficients. Image enhancement using this approach supports low computational complexities, easy modification of the image frequency make-up and easy of pertinence of the transformed domain attributes. Nevertheless, it lacks the ability to explicitly enhance all part of the image concurrently and shows difficulties in automating the enhancement procedure. Some level of experimentation is required to select a particular enhancement technique. It is also pertinent to note that the area of image enhancement in image processing is very subjective.

Some of the work done on image enhancement by various researchers is discussed here. Arun et al. [11] proposed that the Adaptive histogram equalization yields an enhanced
outcome. However, this left the image still bound with faded appearance. The background information and the plane remained clouded, with poor contrast and sharpness. The entire image is reduced to dark tones when an Alpha rooting operation is performed, the outline visible with histogram equalization vanished.

Agaian et al. [3] submitted that the global histogram equalization, that sets out to modify an image spatial histogram resembles a uniform distribution. However, the key limitation in Histogram equalization is the global management of an image which results in poor retention of local image details. Furthermore, the image is over-enhanced leading to loss of visual data in quality and intensity. Tanget al. [112] advocated the use of global histogram equalization, a technique that adjusts the intensity histogram to an estimated uniform distribution. The drawback is that the global image attributes may not be applied suitably in a local setting. Actually, all regions are treated equally in a global histogram alterations, giving rise to poor detail preservation in the local context. Consequently, several other image enhancement algorithms that improve the local image enhancement have been proposed.

UE imaging technique is based on the elastic properties of living tissues in vivo, and the gray scale images that show the distribution of strains in the tissue is known as elastogram. There are two phases involved in forming an elastogram. The first phase involves the exterior compression of the tissue to yield an interior strain field. Acquisition and processing of the ultrasonic echo signals prior and subsequently represents the second phase. The distribution of the elastic modulus and the boundary surroundings influences the compressed tissue strain distribution. The approximation of the strain distribution is dependent on the ultrasonic echo signal collected before and after compression. The signals are segmented at different profundity by partly overlapping windows; the cross-correlation of the waveforms is used to estimate the time delays and subsequently the axial strain fields. The image resolution is improved by applying overlapping windows to the data. The window
length is also a parameter that determines the resolution. A high resolution but noisy or “speckled” UE image is produced by using a small window with a fixed overlap.

When a lucid source and non-lucid sensors are used to probe a medium with a rough wavelength scale, “speckle” is produced giving a granular appearance. Speckle is a type of multiplicative noise that might make it challenging to discern and interpret UE images. Though the speckle in some instances are vital for tagging features, in most instances the speckle noise only deteriorates the quality of the image and damages the edge delineation. The contrast resolution is also lowered, decreasing the ability to detect small and low contrast lesions in the breast image. Filtering out speckle is essential, since it is the primary source of noise in UE imaging.

There are three groups of speckle noise reduction methods: filtering methods [4,102], wavelet domain methods and compound approaches [2, 4, 123]. The most fundamental operation in US/UE breast image enhancement is filtering. Rightly, the term referred to as “filtering” is the measure of the strained image at particular location and is a mapping of the measures of the original image in a small region of the same location. The Gaussian low pass filtering shows good results for medical breast imaging. It exhibits no resounding result as it is a calculation of the weighted average of the pixel measures in the region, in which, the weights declines with length from the regions kernel. Images naturally differ throughout distance, hence neighborhood pixels are expected to have the comparable measures and thus suitable to average them in concert. The noise values that debase these neighborhood pixels are equally less connected than the signal values; consequently noise is removed at the same time preserving the signal.

US and UE imaging demands definite filters because the speckle intensity is dependent on the signal, hence most denoising methods employing the linear Gaussian noise model will not surface. Presented here are some of the speckle reduction methods based on
standard adaptive filters. In US image enhancement the adaptive filters are chiefly used because of the ease of execution and operation. Most of the adaptive filters assume that speckle noise is multiplicative like Kuan filter [63], Frost filter [39] and Lee filter [65]. Improvements of these classic filters have been proposed.

Several research [51, 80] have looked into using schematic wavelet thresholding for speckle reduction, on the premise [35] that the speckle in US image is transformed into a linear Gaussian noise by the logarithm densification. In the wavelet domain image enhancement is achieved using the DWT. Here image are transformed into an estimation sub bands comprising of the scale coefficients and a band of detail sub-band made up of the wavelet coefficient at different scales of orientation and resolution. DWT delivers a suitable foundation for noise separation in an image. Since wavelet transform shows good energy compaction, the lower coefficients will most likely denote noise, and the higher coefficient representing vital image features. The feature coefficients normally continue diagonally and develop into spatially connected bands within specific sub-band. These attributes make DWT appealing for denoising. Several wavelet-based despeckling procedures have been developed. Pizurica et al. [93] in a bid to disprove this premises suggested the wavelet based Generalized Likelihood proportion conceptualization and levied no anterior on noise and signal datum.

Achim et al. [1], and Gupta et al. [48], used the Bayesian model as a basis of using wavelet thresholding conforming to the non-Gaussian information of the signal. Other researchers like Yang [126], and Zhang et al. [128] used multi-scale schemes in improving the performance of the Anisotropic Diffusion (AD) filter.

To improve the Frost and Lee filters, Lopes and Sery [69] proposed pixels classification allowing for specific processing to the different classes. Relying on this notion, further specific areas to be processed can be determined by exploiting the local image statistics using the supposed Adaptive Speckle Reduction (ASR)filter. Karaman et al. [61]
fitted the core of the adaptive filter to the similar region based on the local image statistics. Moreover, Tay et al. [113] developed a stochastic approach to speckle removal in ultrasound images. The basis is the removal of the local extrema accepted as the outliers in a sturdy statistical approximation model. The Rayleigh-Maximum-Likelihood filter was inferred from this theoretical framework [12].

Other approaches to speckle noise reduction based on the Partial Differential Equation (PDE) have been proposed. Perona and Malik [90] developed the AD, while Rudin et al. [97] suggested the Total Variation minimization system for US imaging. Yu et al. [127] introduced the Speckle Reducing Anisotropic Diffusion (SRAD) filter and is based on the variation in the noise-subordinate coefficient.

However in contrast to the former adaptive speckle filters, the PDE filter is iterative in nature, that preserve the edges and develops smooth images. Despite these appealing features, vital structural details are discarded unfortunately in the iteration process.

The different approaches mentioned above can be used in combination to exploit the merits inherent to the different models. Hao et al. [52] exploited this by preprocessing the image using adaptive filters and afterwards decomposed the image into two portions. Consequently the two portions are subjected to Donoho’s soft thresholding. Ultimately, the two processed portions are combined to slash speckle.

Czerwinski et al. [30] proposed the use of “stick”, a linear scheme that resembles a corresponding filter in additive white Gaussian noise. The sticks showed near optimally performance when the speckle is uncorrelated. Remarkable improvement can be achieved in colored speckle, if the pre-whitening steps are taken prior to matched filters and if the exact speckle statistics are known. However, if the speckle statistics are not available, the stick detector still performs optimally in diverse noise environments. Performing a stick operation
exhibits very minute details from the original image plainly, and simultaneously smooths the speckle. This is an advantage for unsupervised image contours detection.

2.4. Feature Extraction

Defining a set of features to be able to describe the density of the breast tissue is an ambitious task for the development of an unsupervised system, particularly if the task at hand is to create an automated approach. It is quite a challenge to examine the elastogram visually and categorize it into one of the five or more types of usual density classification, because occasionally a second reader is required and consequently the automatic algorithm can be crucial in determining the level of density.

One of the vital steps in breast cancer detection and classification are feature extraction and selection. An optimal feature set should significantly discriminate features and effectively reduce redundant feature to circumvent the “curse of dimensionality”. The “curse of dimensionality” suggests that high dimension data may contain to great extent redundant information and reduces the efficiency of the scheme. This is because according to the curse of dimensionality, any increase in dimensionality increases the amount of time required for training data exponentially [38]. In some innovative techniques, like the artificial neural network and support vector machine, the performance of these algorithms in terms of classification and computational time is affected by the dimension of the feature vector. Hence, the most important feature of a good CAD system is the extraction and selection of useful and vital features. Breast UE images features can be divided into categories, but not all can be used at the same time. A necessary step is the extraction and selection of the beneficial features. In selecting significant features four main considerations according to the general guideline should include: independence, reliability, discrimination and optimality, not necessarily in that order. Nevertheless, combination of the best performing feature is not an
assurance of an efficient system. The aim of feature extraction and selection is to exploit effectively the discriminating operation of the feature group.

The texture, edge and wavelet feature of the distinct region may be used to differentiate between healthy and unhealthy tissues. This can be represented by statistical descriptors extracted from the co-occurrence matrix, the wavelet transforms and Local Binary Pattern (LBP) besides structural and spectral descriptors. A combination of various set of features has been used in order to obtain the best breast tissue density representation.

2.4.1. Wavelet Features

The wavelet transform is a mathematical tool with a great variety of applications, and one important example, is signal and image analysis. By a wavelet transform data is split into different scaling components and then each component is analyzed with a resolution matched to its scale. There are many types of wavelet families, for example functions suitable for continuous transformations and the orthonormal set of functions with compact support, which define an important group of discrete wavelet transforms, spline -wavelets and many others. There are several wavelet families, such as the functions presented by Meyer [79], which are apt for continuous transformations. Another commonly used wavelet is the orthonormal group of functions with compressed support introduced by Daubechies [31]. They are parallel or analog transmutations that function on a data vector that has a length of a whole number with the power of two, translating it into dissimilar vectors of similar length L2(R). This contains the finite signals explored in this research. It is a potent instrument that can be used to distinguish data into different frequency elements, and then performing a clear cut analysis of each component in correspondence with its scale. Daubechies states that the wavelet family is linked to filter bank methods in digital signal processing, and hence useful and proper for discrete and fast transformations. The equal ratio relationship between Daubechies
wavelets and finite sized filters gives this family an important edge in the study of discrete signals [116].

Algorithms based on wavelet transform have been used in various unsupervised detections of microcalcification. Zadeh et al. [109] were able to show that the multi-wavelet approach has a superior performance in their study of 103 delineated areas with microcalcification. They set out by extracting three different types of features from the Region of Interest (ROI): co-occurrence, wavelet and multi-wavelet based features and used the K-Nearest Neighbor (KNN) to categorize benign and malignant clusters. The multi-wavelet technique gave the best result of 0.91 for the area under the Receiver Operating Characteristics (ROC) curve.

Multi-scale depiction of function is one of the main advantages of DWT. It easily allows for analysis of functions at different degrees of resolution. It is also invertible, allowing for image reconstruction from its discrete wavelet coefficients. Wavelet transform can be reused after prior analysis and the signal can be reconstructed without loss of information. Other key attributes include regularity and translation covariance [99]. These improve sensing of distinctiveness and allow for original image shifting which creates a wavelet coefficient shifting with no change to its structure.

2.4.2. Texture Features

A good texture descriptor must effectively represent all desired features characteristics and must be constant in viewpoint. The linear or nonlinear transformation of the coordinate scheme is known as feature extraction. The principal component analysis (PCA) is a renowned feature extraction method. PCA executes on the covariance matrix symmetry, and resolves the characteristic root of a square matrix and eigen vector. PCA system performs very well in the reduction of high dimensional interconnected features into
low dimensional features. PCA can effectively optimize the feature vector of the auto-
covariance coefficients [60]. The discriminant analysis, independent component analysis and
the factor analysis are other feature extraction methods that can be utilized in dimensionality
reduction.

A simplistic approach to texture description is to use the statistical moments of the
intensity histogram of an image or region [43]. The drawback of using only the histograms in
calculation is that texture measures will only represent information about distribution of
intensities, but nothing on the position of pixels with respect to one another within that
texture. Employing the co-occurrence matrix substantially solves the problem of information
regarding comparative position of the neighboring pixels in an image.

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. It was clearly described by Haralick et al. [53] nearly four decades ago. 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 [110].

The matrix is a construct of a distance d and direction of θ (given as 0°, 45°, 90° and
135°). Textural features are usually extracted using GLCM contrast, correlation, energy,
entropy, local homogeneity, and shade [8].

Several studies have described the extraction of useful textural features for tumor
detection based on the gray-level co-occurrence matrix (GLCM) [9,82]. Tsiaparas et al. [115]
proposed the extraction of textures in wavelet domains with a multi-scale approach and
Support Vector Machine classifier. Gómez et al. [42] analyzed the gray-level co-occurrence
statistics with six quantization levels and proposed the use of the selected effective texture
descriptors for US breast tumor diagnosis.
Extracted first-order statistics as well as second-order statistics (like gray level co-
ocurrence matrices) were used by Landeweerd and Gelsema [64] to discern types of white
blood cells. Chen et al. [26] used an enhancement algorithm based on multi-scale wavelet
analysis to extract and review information about each scale. The development of wavelet
coefficients on different measures was able to define the shape of uneven bodily structures.
This construct can be applied to breast tissue images to enhance composite bodily structures
and elusive tissues. Fractal texture features are used to classify US images of liver and also
used the fractal texture features to enhance the edge in chest X-rays. Additionally, useful
texture features are extracted from wavelet transformed US breast images and examined.

Lundervold [70] employed the use of fractal texture features like reception to edge
operators to study the heart US images. In the study the images are time sequence of the left
ventricle. At each pixel texture is denoted as an index, this is similar to the local fractal
dimension between an 11 x 11 window calculated using the Brownian motion model
proposed by Lin et al. [66]. This texture feature is used in addition to other conventional
features, with the gray level, reception to Kirsch edge operators and the outcome of time-
based processes.

It is assumed that the fractal dimension will be more prominent on a typical blood
vessel than in other tissue because of dispersions and noise. Moreover, the fractal dimension
is depleted at the purposive blood and tissue port constituting the edge information.

However, a difficult task for orthodox texture analysis in US breast images has been
discerned, notably the disjointed texture feature extraction from the original images or
wavelet domain images which subsequently results in the challenges of variant texture
analysis in tumor detection. Taking into consideration the high demand for invariant texture
analysis in developing clinical applications, research is directed towards developing methods
of extracting gray-scale invariant texture features for shape identification.
The Local Binary Pattern (LBP) [86] was first studied in gray-scale and rotation invariant texture classification. LBP operator has the features of both statistical and structural texture analysis. LBP operators are mainly used in gray scale invariant two-dimensional (2-D) texture analysis. In LBP operators the pixels of an image are marked by thresholding the neighboring (i.e. $3 \times 3$) pixel with the center value and counting the result of the thresholding as a binary number [68].

After marking all the pixels with the equivalent LBP codes, computed histograms of the markings are then used as texture descriptors. Masumoto et al. [78] proposed a method of extracting the textures based on LBPs for separating solid masses in US breast images. The outstanding properties of LBP features are computational simplicity and permissiveness opposed to monotonic brightness variations [56].

2.4.3. Edge Features

The process of describing and finding sharp incoherence in an image is known as edge detection. Incoherence’s are sharp variations in pixel intensity which are common features of boundaries of objects in an image. Several edge detection operators are available (Prewitt etal. [94], Sobel [107]), and are contrived to be responsive to some particular kinds of edges. Edge detection has largely been used for object tracking, segmentation etc.

For decades various algorithms has been proposed for extraction of homogenous regions in digital images. Edge detection algorithms have gained a lot of prominence. Definitively, as found in the classical edge detection techniques (Robert, Sobel, Prewitt and gradient operators) they evaluate the image intensity derivatives. The use of surface fitting and smoothing filter employed as regularization methods is to counter the effects of noise on differentiation. Zhang and Bi[129] suggested the use of Gaussian pre-convolution for other types of noise. Nevertheless, the use of Gaussian-like smoothing is not without limitations, as
while the noise are smoothen out, high frequency edge features are also removed, reducing
the ability to detect low contrast edges. Classical edge detectors focuses on the high
frequency constituents in the image and fails to account for the moderate and low spatial
resolution in an imaging device, and hence, perform poorly in these cases.

The recognition of this limitation gave rise to new techniques that tends to deals with
suppression of noise, blurring of image and also the capability of effectively resolving edges
interference. Band-pass filters like the Canny operator fits perfectly into these techniques.
The most commonly used edge detection algorithms such as; Canny, Kirch, Laplacian and
Marr-Hildreth were appraised and classified by Sharifi et al. [103]. An evaluation by Shin et
al. [105], showed the Canny detectors as having a very good test performance with
remarkable robustness in convergence, and has the lowest computational time among other
detectors. This is consistent with result showed by Heath et al. [54].

A new algorithm for edge detection in computed tomography (CT) medical images of
the lungs was proposed by Zhao YQ et al. [130]; this algorithm was shown to be most
effective for edge detection and medical image denoising than other edge detection algorithm
like Sobel edge detectors and Laplacian of Gaussian operators. A new edge detection
algorithm founded on a statistical method employing the student t-test was proposed by
Fesharaki and Hellestrand [37]. They proposed a 5x5 window partitioning in eight different
orientations for edge detection. With one of the partitioning corresponding to the direction of
the edge in the image with the highest values for the statistic defined in the algorithm. This
method was shown to meaningfully suppress noise and at the same time preserves the edges
without prior information about the noise in the image.

Homogenous neighborhood detection in the edge function derived from the multi-
resolution edge sensing is useful in edge detection. The goal of edge sensing is to create an
dge function, based on the dissemination of the intensity disjointedness of the image; the
region can be defined as homogenous if it has a very small edginess portion. Edge direction and homogeneity intensity has been used as a method for edge detection; this was effectively employed by Boland et al.[19]. For UE gray scale images, the gray level disjointedness is minute between healthy and unhealthy tissues. The edge gained from the edge detector is a derivative of the diverse scales showing the strengths and precision of the edges. Hence, a smaller scale gives finer and more accurate edges. Nonetheless, a small scale edge detector is sensitive to noise.

2.5. Feature Selection

One problem inherent in high dimension data is that it may contain redundant information to a great extent and reduces the efficiency of the scheme. This is because according the curse of dimensionality, any increase in dimensionality increases the amount of time required for training data exponentially. Likewise, there is a possibility of a strong correlation among diverse feature. And this serves as a motivator for feature size reduction. Feature reduction can generally be achieved by (a) selection of the best subset of features that makes the most of the objective function or (b) by feature transformation into a new space and reconstructing the new as a combination of the original feature. The objective function can be defined using the predictive accuracy of the subsets. This was done by Baheerathm et al. using the leave-one-out approach on the training set to estimate the categorization accuracy. Objective function can also be defined as the estimation of the subsets in order to apply their data content. The data content is evaluated by employing the interclass space values like the Euclidean and Mahalanobis values[6] or the feature dependency on the data statistics [14]. In the second instance the features selected in the subset are the feature that exhibits meaningful diversity between the different classes. Feature selection can be performed using exhaustive search or using sequential search algorithm. The exhaustive
search involves the testing of all possible features subsets. But, this search is impractical as the possible subsets of features grows exponentially even for average number of features. However, the size and content of the subsets can be limited to control this problem.

Sequential search algorithms works by removing or adding features sequentially. Sequential floating selection, sequential forward selection, sequential backward selection, and plus-l minus-r selection are examples of sequential search algorithm. Walker et al. [117] and Weyn et al. [119] used the plus-2 minus-1 selection and sequential floating selection procedures respectively in reducing feature set sizes in an unsupervised cancer detection system. Sequential forward selection sets out with a null set of features and additively grows the subset by feature addition. These added features are carefully chosen so that the successive subsets give rise to the best objective function. The search ends when no additional features leads to a better objective function.

Similarly, sequential backward search sets out with a complete set of features and additively get rid of the inoperable features one after another. Nevertheless, the two search techniques are representative of the greedy algorithms examples and unfortunately do not assure the best potential solution and feature selection are based on the assumption that the feature are independent.

Though, some features might not produce a desirable result when used solely, and the best objective function are only realized when the feature are used in combination with other features. Taking into consideration that, the sequential forward and backward selections works additively it fails to appropriate such complimentary features. To get around this difficulty, Pudil et al. [95] proposed the plus-l minus-r selection. This algorithm sets out with a null set and incrementally adds and subtracts features (adds 1 feature and subtract r of them) in each step. The challenge is then in selecting the optimal measures of l and r. Instead of using fixed values of l and r, the sequential floating selection permits the use of different
measures of 1 and r each point. The Sequential Floating Forward Selection (SFFS) feature selection approach is used in handling this kind of nesting problem. These selection algorithms all have prospect to be rein in local minima, and therefore, providing suboptimal solution.