International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2020 IJSRCSEIT | Volume 6 | Issue 3 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT20633

Breast Cancer Detection via Mammographic Images: A Survey

Mary Walowe Mwadulo[1*], Raphael Angulu[2], Stephen Makau Mutua[3]

[1,2]Department of Information Technology, School of Computing and Informatics, Masinde Muliro University of Science and Technology P.O. Box, 190 - 50100, Kakamega, Kenya

[3] Department of Information Technology, School of Computing and Informatics, Meru University of Science and Technology P.O. Box, 60200 - 972, Meru, Kenya

*Corresponding Author: mwalowe@yahoo.com

ABSTRACT

Breast cancer is a top killer disease for women globally. The long term survival rate of women can be improved through early and effective screening of breast cancer cells. Currently, a mammogram is the recommended tool for breast cancer screening since it can identify breast cancer cells several years before physical signs appear and it is cost effective. This paper analyzes mammographic detection of breast cancer by providing an explanation on development and classification of Breast Cancer, Image representation models for breast tumor, mammography technologies, a discussion on various mammographic signs of breast cancer, breast cancer feature extraction techniques, popular breast cancer classification techniques, comparative analysis of existing mammogram breast cancer databases, and a review of mammographic breast cancer detection studies are presented. Finally, a highlight on future work is given.

Keywords : Breast Cancer, Mammography, Feature extraction, Classification, Database

I. INTRODUCTION

Cancer is a large group of related diseases caused by an uncontrolled division of the body cells that spread, crowd out normal cells and develop into a tumor that is either benign or malignant [1]. Cancer can begin anywhere in the human body as such, the name of the cancer is dependent on the affected area. Among the different types of cancers, breast cancer is a leading cause of women's deaths.

Breast cancer occurs due to abnormal growth of breast cells. The cancer cells divide faster than healthy cells, accumulate and form a mass. Breast cancer can either begin within the cells of the lobules that are the milk-producing glands or the ducts, that drain milk from the lobules to the nipple. Though unlikely, breast cancer can even begin within the stromal tissues, that embody the fatty and fibrous connective tissues of the breast[2]. Gradually, cancer cells can spread to nearby tissues such as lymph nodes and find their way to other body parts. Although breast cancer can develop in men, the highest risk and incidence is in women especially those who are 50-64years of age [3][4] [5][6][7].

Breast cancer is rated as the second most frequent cancer killer for women in the world [8]. In the year 2018, an estimated 2,088,849 (11.6%) new breast cancer cases were diagnosed and 626,679 (6.6%) breast cancer deaths were reported [9]. Its survival rates vary greatly in the world because it is dependent on factors

such as age, geographical factors, and race. However, a relative survival estimate is 91% at 5 years diagnosis, 86% after 10 years and 80% after 15 years[10]. A report by IAR on Cancer indicates that there are more deaths in less developed regions than the more developed regions, partly because a shift in lifestyle is causing an increase in incidences and clinical advances to combat the disease are expensive and sometimes unavailable.

A research conducted by Korir et al [11] between 2004 to 2008 on cancer incident rate among Nairobi women, discovered that breast cancer was rated highest accounting for 23% of all the cancer cases. Early breast cancer detection is important because it increases the survival rate for women diagnosed with the disease and contributes positively to better health care.

Currently, breast cancer detection has gained significant attention with most studies focusing on ways to detect breast cancer in its early stages using advanced technology. Latest Studies in [12][13][14][15] show that even though techniques have been developed to combat the disease, there is a need for better methods for handling the disease.

II. DEVELOPMENT AND CLASSIFICATION OF BREAST CANCER

Breast cells contain Deoxyribonucleic acid (DNA) which instructs the cell on how to look and behave. Each time a new cell is formed by division, the instructions (DNA) are copied. Normally when the division occurs, a number of checks happen to make sure that the DNA has been copied correctly[16]. However, sometimes they are not correctly copied. Even though DNA contains repair genes that are able to self-correct the mistake, sometimes the mistake (mutation) does not get corrected; instead, it is passed on. Over time, mutations accumulate forming a tumor that can be cancerous. The more often a cell divides, the greater the risk of mutation occurring and accumulating. Anything that accelerates the rate of cell division also increases the likelihood of mutation occurring[16]. For instance, a high concentration of estrogen in a female breast stimulates cell division giving the DNA less time for repair leading to DNA damage and mutation[17]. Other agents such as X-ray, Ultraviolet and some chemical exposure can increase mutation rate by damaging DNA directly.

The type of breast cancer is determined by the specific cells in the breast affected. Most breast cancer affects the breast Duct and Lobule. Extent of breast cancer is described as either in situ (has not spread) or invasive (has spread to surrounding tissues). The in situ and invasive have characteristic patterns which are used to classify them. In situ is divided into ductal and lobular which originate from ducts and lobular and do not destroy other tissues[18]. Ductal carcinoma in situ (DCI) is characterized by the proliferation of malignant mammary ductal epithelial cells that line the breast milk ducts. It is non-invasive as it is confined to the breast ducts however it accounts for 90% of breast cancer [19]. When detected early, DCI is highly treatable, but if left untreated it can spread to surrounding tissues.

Lobular carcinoma in situ (LCIS) occurs before menopause and originates from milk glands. This type of carcinoma is challenging to detect in mammography and therefore between 25 and 20% of women presenting this kind of tumor develop invasive breast cancer. Invasive Ductal Carcinoma (IDC) starts in the milk-producing ducts and penetrates the wall of the duct, invading the fatty tissues of the breast. IDC accounts for 80% of the breast diagnosis [20].

The extent of breast cancer is described in four stages. The definition of a stage is based on the size of the tumor, whether it is invasive or non-invasive, whether it is in the lymph nodes or not and whether it has spread to other parts of the body beyond the breast. Stage 0 is the earliest stage in which the cancer has not

spread while stage IV the cancer has spread to distant organs or the lymph nodes far from the breast.

III. IMAGE REPRESENTATION MODELS FOR BREAST TUMOR

Locating and correctly interpreting a breast tumor is a viable though challenging task in computer vision because of the inherent variability of the tools used and breast structure. While tools used to capture the breast image are susceptible to noise and illumination, breast structure among individuals varies. While some individuals may have fatty breast others have dense breasts. This section presents various approaches discussed in the literature for image representation when detecting breast cancer cells.

A. Shape model

Shape is a significant visual cue used by radiologists to distinguish the two different classes of breast tumors. Benign tumors have circular and symmetric shape while malignant tumors have random and asymmetric shapes [21]. To identify the shape of the tumor, the radiologist uses the shape interior region or contours defined by the tumor boundary. An effective shape model should successfully discover similar shapes in a pattern matching problem, irrespective of whether the shape being recognized is rotated or flipped. Also, it should be robust such that it can effectively compare two images even though one may be noise affected or distorted.

When the shape interior region is used, all the pixels within the shape are taken into consideration for shape representation. Further, the interior region can be considered as global (takes the entire region into consideration) or structural (Partition the interior region into sub-parts) for shape image representation[22]. Shape contours are defined by tumor boundaries. The boundary defines the edges which define a transition between tumor and surrounding tissues. Understanding tumor boundary aid in knowing to what extent the cancer cells have spread, which brings out the size of the tumor. The shape contours can be represented globally in which the complete margin information is used to develop the feature vector or structural in which the shape boundary information is broken into sub parts[22].

Shape features such as roundness, eccentricity, and compactness are used to characterize a tumor[23]. Caulkin et al [24] defined shape vector $x_{shape} = (x_i \ldots x_n, y_i \ldots y_n)$ by placing k equally spaced points on margins of each mass then defined the origin as the point of intersection between the border and the line connecting the nipple. This approach was based on the understanding that masses are formed in the breast ducts that originate from the nipple. However, this introduced rotational variances in the model, because the model did not consider interdependencies between size and feature vector. To improve the model Berks et al [25] defined a compact model of shape variation built by applying PCA to the aligned vectors Such that $x_{shape} \approx x'_{shape} + P_{shape}\, b_{shape,}$ where p_{shape} is a matrix of principal modes and $b_{shape,} = p'_{shape}\,(x_{shape} - x'_{shape})$ is the vector of model parameters stored for each shape. This model had negligible variance discarded.

An important characteristic of using shape to locate breast tumors is low computation complexity and robustness of shape features [22]. However, using shape features requires a sufficient number of landmark points to reveal the complete shape of the tumor.

B. Appearance model

Texture models, mainly model the presence of a tumor by looking at structural properties which are not visible to the human eye. They have been used extensively in modeling tumors for breast cancer detection. They take into consideration the texture of a tumor which makes the model appropriate for breast cancer detection at all stages. Texture is a feature of homogeneity of images using the pixel for tonal

variation, which has a certain scale, regularity, and directionality [26]. Texture highlights important details about the structural arrangement and environmental relationship of the object in an image and reveals important discriminatory characteristics related to variability patterns[27]. While shape models are inspired by the properties for which a radiologist looks for, appearance models are structured in a way that they can capture important properties not noticeable to the human eye[28]. Therefore, appearance models can capture properties that are valuable but not extracted visually. Also, they do not occur over a point but rather over a region. Even though appearance models are used to analyze many images in natural and medical science, extracting the features is challenged by changes that could result from illumination, position view, which cause large variation in the image appearance of texture. Also, two different textures under different imaging conditions can appear to be similar[29]. Furthermore, to effectively use appearance models there is a need to first process each region to remove background breast tissue[25].

C. Hybrid models

Hybrid modeling seeks to improve on the limitations of shape and texture models by combining them with the goal of obtaining a more robust model for breast cancer detection. Hybrid modeling aims at improving the detection of masses and microcalcification by increasing the rate of true positive and reducing the rate of false-positive, in so doing the CAD system is improved[23]. In hybrid modeling, the performance is expected to be higher than when using either shape or appearance models. However, care should be taken to ensure no two or more similar models are combined as no much improvement will be achieved. Also while modeling a hybrid model care should be taken such that no two or more combined models cancel the effects of each other. Models are combined either in parallel, in which individual outputs of the models are concatenated and used for final breast cancer detection or hierarchical in which output of the first model is fed as input for the second model. The purpose of this approach is to take advantage of the strength in each of the modeling techniques.

IV. MAMMOGRAPHY TECHNOLOGIES

There are two types of mammographic technologies: Screen-Film Mammography (SFM) and Full-Field Digital Mammography (FFDM) [30] also called digital mammography. The two types of mammography use x-rays to develop an image of the breast, however, they differ in the way an image is taken. SFM forms a breast image on photographic films while FFDM takes an electronic image on digital files which are directly recorded on a computer[31],[32]. SFM has advantages in terms of cost and availability over FFDM, however, they suffer from limited dynamic range of exposure, the image can only be stored on the film and once an image is created, its contrast cannot be altered, the film serves as a detector, display and archive medium [33]. Furthermore, SFM is not appropriate in women with radiographically dense breast tissue because of low sensitivity [32]. FFDM, on the other hand, has the ability to store the image electronically, it is less susceptible to noise and image quality can be enhanced to make evaluation of the x-ray a lot easier, thus producing more accurate results [34]. Even though FFDM has more advantages than SFM, clinical tests in [35][3][36][37] showed that in terms of overall detection accuracy, there is no significant difference between FFDM and SFM. However, for women between 50-64 years of age, FFDM resulted in higher detection and recall rates than SFM [3][4] [5][6][7]. Furthermore, FFDM gives more accurate results than SFM for pre and perimenopausal women with dense breast tissues younger than 50 years [31][35] [38].

Even though, FFDM has been recommended as an effective screening [39]. However, the reading and interpretation of is done by a radiologist who is susceptible to human observer variability and the

correctness of the results is dependent on the skills of the radiologist administering and reading the FFDM. Moreover, dust particles on an FFDM and breast surgery scars on a patient could obstruct the radiologist and lead to false interpretation[40]. This makes FFDM susceptible to a high rate of false-positive and false-negative. As a measure towards improving the effectiveness of the mammogram results some radiologists resulted to double reading of mammograms. This is not a feasible solution because it is economically costly and time-consuming [41][42]. Therefore, the adoption of a CAD system is a viable solution since it can automatically enhance the contrast between the region of interest and its background which allows differentiation between abnormal and healthy tissues. Secondly, CAD systems can extract significant discriminant features, then learn based on the features extracted and predict absence or presence of cancer tissues independent of the radiologist. Finally, the preprocessing stage in CAD system can eliminate noise and artifacts found in mammographic images.

V. MAMMOGRAPHIC SIGNS OF BREAST CANCER

Microcalcification, Mass, Architectural Distortions, and Bilateral Symmetry are four signs of breast cancer cells found on a mammogram. Mass and Microcalcification are the most common indicators. While masses are commonly seen among patients dragonized with invasive breast cancer, microcalcification is reported in a higher percentage among patients diagnosed with Ductal Carcinoma In Situ (DCIS breast cancer. Bilateral Asymmetry and Architectural distortion breast cancer signs are rarely linked to the presence of breast cancer cells. A research conducted by Venkatesan et al [43] on 1552 breast cancer cases where 1287 were invasive breast cancer, masses were more present at 56%, followed by microcalcification with 29%, Bilateral asymmetry with 12% and 4% for architectural distortion. Gadjos et al [44] concludes that 95% of breast cancer presenting as masses were invasive cancer while 68% of breast cancer presenting as microcalcification was associated with DCIS.

A. Micro calcification

Calcification is an accumulation of calcium in the breast which can either be macro calcification or microcalcification. Macro calcification is large calcium deposits considered noncancerous therefore, they are not linked to breast cancer and consequently, no special consideration is dedicated to them [45]. Microcalcifications are clusters of calcium deposits that are small bright white dots of varying sizes and shapes in the breast tissue [46]. Benign microcalcifications have a regular shape and are found in isolation while malignant microcalcification has an irregular shape and is clustered[47]. Microcalcification is identified by their size, shape, number, and distribution[28][48]. The more, bigger, and closely clustered they are, the higher the chance is for them to be identified. Microcalcification is a challenge to discover because they are not separated from the surrounding normal tissue. When malignant cells start to invade the tissues the microcalcification viewed on a mammogram will be seen as a light patch on normal tissue.

Developing a model to detect microcalcification is easier because their presences is depicted by their numbers and how they are distributed, however their small size, presences of overlapping breast tissue, low contrast and breast density especially in young women increase the probability that they can be missed or misinterpreted [28],[49]. Studies in [50][51] [52][53] [54] developed wavelet transform-based methods, for detecting microcalcification. The main reason being that wavelet transform can specifically locate the image region with high spatial frequencies than transforms like Fourier than give the content of frequencies but cannot locate in the image the specific

spatial frequencies [28]. In addition, because microcalcification is seen as bright white dots on the mammogram, the wavelet transform methods can easily identify them as discontinued points. Studies in [55][56] opted to detect microcalcification through shape features, though they achieved good results it is clear that shape features alone cannot be used and it is a challenge to detect microcalcification based on shape because of their small size. Other Studies in [57][58] opted to use texture features , but even with good classifiers like SVM and ANN the results were not good enough.

B. Masses

A mass is a lump that varies in shape and size and can be seen in two projection; shape and margin properties[21]. A mass that is round, smooth and has circumscribed margins has an increased possibility to be benign whereas a mass that is spiculated, rough and has blurry margins has an increased likelihood to be malignant [21]. A majority of algorithms used for mass detection have two stages: (1) detecting suspicious regions (2) classifying suspicious regions.

When a suspicious region is correctly identified, it is expected that the sensitivity to correctly classify the region during the classification stage is increased. However, sometimes it results in a high number of false positives. Algorithms applied for stage one detection are of two forms; Pixel-based detection methods and region-based detection methods [59].

Pixel-based detection is based on the extracted features of its local neighborhood, which is achieved by either defining a threshold value or using a sophisticated classification technique [59]. The suspicious pixels which have been detected are then grouped together into regions by connected pixels. Regions labeled as suspicious by the algorithm do not automatically indicate malignancies. Categorization of suspicious region into malignancy or benign id done during the classification stage.

Researchers have mainly focused on pixel-based methods. Liu et al [60] designed a multiresolution system to detect spiculated lesions in mammograms based on a binary tree classifier. Experimental evaluation using MIAS dataset showed the scheme achieved a low positive rate. Sampat and Bovik [61] presented a two-stage algorithm for detecting spiculated lesions in mammograms. The lesions were first enhanced then their location was detected. The algorithm was tested on the DDSM database and results achieved indicate that the algorithm could correctly locate the mass region. Zwiggelaar et al [62] developed a model for detecting spiculated lesions based on local scale oriented signatures built using recursive median filtering and Principle Component Analysis. They achieved a sensitivity value of 80%. Even though a majority of researchers have focused on spiculated masses because of their high probability of showing malignancy, Other researchers such as [14],[63],[64],[65] and [66], focused on masses without considering a specific type of margin.

An advantage of pixel-based methods is accessibility to numerous number pixels per image for training a classifier. However, the ability to use multiple pixels could pose a challenge. When multiple pixels are present on the margin and center of a mass, it becomes a challenge for the classifier to differentiate between these two regions and as such, they are group together to the same class, yet they may not always be homogenous[59]. These pose a challenge of discriminating against a mass from normal surrounding tissue.

In region-based detection methods, filtering or segmentation techniques are employed to obtain the Region Of Interest (ROI). Shape and texture features are then extracted from each ROI and then they are classified as either suspicious or not. An advantage of region-based detection methods is taking into consideration spatial information which improves its

discriminate power of a mass from its surrounding tissues. Also, the features extracted directly correlate with the shape and margins of the extracted regions. These ensure regions with similar characteristics can be categorized together. Further, in comparison to pixel-based methods, region-based are less computationally expensive. However, region-based detection methods have few samples for training if a classifier is to be used.

The purpose of this stage is to classify suspicious region as normal/abnormal tissue and decrease the number of false positives generated in stage one. Olivera et al [67] used the Support Vector Machine classifier to differentiate between normal and abnormal breast regions. The proposed methodology attained an average accuracy of 98.88% with the DDSM dataset. To develop a system that emulates image features used by a radiologist Brake et al [68] relied on intensity, density, location, and contrast features to obtain approximately 75% accuracy of all detected cancer. Wang et al [69] relied on size, shape, contrast, homogeneity and speculation features. They obtained AUC of 0.786±0.026.

Even though studies in [70], [71], [72], [73], [74] established various approaches for identification of masses, it should be noted that detecting a mass is more challenging than detecting microcalcification, because masses have poor image contrast [75]. As such, it is important to properly segment the tumor from its background so as to increase its visibility. Also because a benign and a malignant tumor develop form one spot and spread outwardly, the shape of the tumor is relatively not specific. Therefore, differentiating the shape of a benign tumor from a malignant tumor is a challenging task.

Most 80 -85% of breast cancer cells can be seen as masses, clusters of microcalcification or a combination of both [76]. However, 15-20% of breast cancer cells may not exist as a malignant mass or a malignant microcalcification. In such circumstances, focal nondescript abnormalities that include bilateral asymmetry and architectural distortion may be the only clue that a malignant tumor is present.

C. Architectural distortion

Architectural Distortion (AD) occurs when the normal breast architecture is deformed without the presence of a defined mass. The distortion often appears star-shaped. Even though (AD) are less predominately than masses and microcalcification, they can predict breast cancer with high accuracy at the screening and diagnostic stage [43]. Approximately, 12% - 45% of missed breast cancer cases in mammogram screenings are as a result of the breast cancer cells manifesting themselves inform of breast Architectural Distortion [77]. Consequently, it is significantly important to accurately identify ADs as a measure of reducing the number of missed breast cancer incidences. Even though a high percentage of DCIS is commonly manifested as malignant appearing microcalcification, Architectural distortion accounts for 10.8% [78]. However, AD's is predominantly associated with invasive ductal and lobular breast cancer.

D. Bilateral asymmetry

Bilateral Asymmetry is an indicator used by radiologists to identify the existence of breast cancer cells viewed on a mammogram. To detect Bilateral Asymmetry (BA), the right and left breasts are compared against each other for any inequality. Examining the breast asymmetry can provide insight into symptoms like parenchymal distortion and asymmetric points which cannot be evaluated by other methods[79]. Before comparison, one of the breasts is flipped so that both breasts are on the same orientation. On a computer monitor, different colors are used to provide a visual distinction between regions on the left and regions on the right breast. The regions on the right breast that are different from the corresponding regions on the left are highlighted

using color green while regions on the left breast the are different from corresponding regions on the right are highlighted using color blue [45].

Microcalcification, masses, AD's and BA are all breast cancer signs viewed on mammography. However, their performance sensitivity, specificity, and accuracy vary. A research carried out by [80] revealed that mass detection has the highest sensitivity of 94.7% followed by microcalcification with 93.7%. The sensitivity was good in both microcalcification and masses but poor for Architectural distortion and Bilateral Asymmetry. With regard to specificity values, Architectural Distortion leads with 79.1% and Bilateral Asymmetry being the least with 52.4%.In terms of performance accuracy, masses had the highest with 84.8% followed closely by microcalcification with 82.1%. Bilateral Asymmetry has the least accuracy of 67.4%. The good performance of masses and microcalcification makes them the most popularly used mammographic signs for breast cancer detection. Even though many publications have focused on detecting and analyzing microcalcification and masses, very few researchers focused on detecting AD in mammograms. Broeders et al [81] and Rangayyan et al [82] suggested that a more effective breast cancer prognosis could be realized if more attention is geared towards the detection of AD's.

VI. BREAST CANCER LOCAL TEXTURE FEATURE EXTRACTION TECHNIQUES

Texture is a feature of homogeneity of images using the pixel as the fundamental for tonal variation, which has a certain scale, regularity, and directionality [26]. Texture highlights the structural arrangement and environmental relationship of the object in an image and reveals important discriminatory characteristics related to variability patterns[27]. An explanation of local texture feature extraction techniques applied for breast cancer detection is presented below.

A. Gabor filters

The Gabor filter is a texture analysis technique that uses different frequencies and orientations modulated by a gaussian function[83]. A 2D Gabor filter is computed as;

$$g(a,b) = \left(\frac{1}{2\pi\sigma^a\sigma^b}\right) exp\left[-\frac{1}{2}\left(\frac{a^2}{\sigma_a^2} + \frac{b^2}{\sigma_b^2}\right) + 2\pi jSx\right] \quad (1)$$

In which σ_a and σ_b are standard deviations of the distribution while S is the radial frequency.

The general equation for a Gabor filter bank is:

$$g_f(a,b) = z^{-m} g(\bar{a},\bar{b}) \qquad (2)$$

Where,

$$\bar{a} = acos\theta + bsin\theta and\ \bar{b} = -asin\theta + bcos\theta$$
$$\theta k = \pi\frac{k-1}{n}, k = 1,2,3 \ldots . n$$

where n is the number of orientations used and z^{-m} =0,1,2…S for S scales.

The 2D bandpass filter has demonstrated optimal localization properties in both spatial and frequency domains. Hence, making it suitable for extracting image features positioned in a particular orientation and within a certain frequency. Since breast cancer masses are defined by shape and margin in a given orientation. Gabor filter has been extensively been used for mass segmentation and edge detection in breast cancer mass detection. Salabat et al [84] Initialized different filters in Gabor bank to different scale and orientations for the purpose of extracting patterns in ROI to help differentiate between normal and abnormal breast tissues.

B. Linear Discriminant Analysis

The Linear Discriminant Analysis (LDA) is a dimensionality reduction strategy that aims at transforming a high dimensional feature vector to a low dimensional space by increasing the ratio between intraclass scatter S_w and inter-class scatter S_b.

S_w and S_b measures are given by:

$$S_w = \sum_{b=1}^{c} \sum_{a=1}^{N_b} \left(x_a^b - \mu\right)\left(x_a^b - \mu\right)^T \quad (1)$$

Where x_a^b and μb ae the a[th] instance and mean class of b respectively, while b,c is the number of classes. N_b is the number of instances in class b

$$S_b = \sum_{b=1}^{c} (\mu_b - \mu)\,(\mu_b - \mu)^T \quad (2)$$

C. Gray level Co-occurrence Matrix

The Gray-level co-occurrence matrix is a texture measure that utilizes image distance and orientation to examine the texture of a grayscale image in which every pixel is compared with its neighboring pixel. The image distance between the reference pixel value and its neighboring pixel values forms a square shape [12] quantized in 0^0, 45^0,90^0,180^0 orientation [71]. Also, it is balanced about the diagonal, such that if there is a difference of 2 cells from the diagonal, then it is a two-level gray difference[85].

The GLCM value at GLCM (x,y) is defined by probability measure of reference pixel x and gray value y at neighboring pixel by distance d, orientation θ given by :

$$P_{d,\theta} = P(x, y) \quad (1)$$

Where $P(x, y)$ is defined by

$$P(x, y) = \frac{GLCM\ (x, y)}{\sum_{x=0}^{N} \sum_{y=0}^{N} GLCM\ (x, y)} \quad (2)$$

Even though GLCM is easy to implement, has good performance in terms of processing time and complexity [86] and gives good results in large field of application, it however suffers from having large dimensionality which forces it to reduce the number of gray levels and may not be effective in images with a lot of noise. Also, the image quantization process leads to loss of information, making the extracted features not reliable[28]. Lastly, there is no established way of choosing the displacement vector (d) and calculating co-occurrences matrices for different values [28].

D. Local binary patterns

The Local Binary Pattern (LBP) is a local texture descriptor proposed by Ojala et al [87]. Given an N × N image, LBP operator thresholds p neighboring pixels with the central pixel results to an 8-bit binary code. The LBP operator considers p neighboring pixels along a circular path either clockwise or counter-clockwise and R distance which is radius of comparison as shown in the formula below:

$$s(x) = \left(x_c, y_c\right) = \sum_{i=0}^{7} s\ (g_i - g_c)2^i \quad (1)$$

$$s(x) = \begin{cases} 1 & if\ x \geq 0 \\ 0 & if\ x < 0 \end{cases} \quad (2)$$

Where g_i the value of its neighbors is, $g_c$is the gray value of the referenced pixel, x_c is the total number of neighbors and $y_c$is a neighborhood radius. First, the center pixel value is identified then it is compared with the neighbor values using a defined radius. The LBP extraction algorithm is defined by two main steps: thresholding and the encoding step.

In thresholding, a central pixel is identified and all the neighboring pixel values P based on an identified distance R are compared with the value of the center pixel. If the neighbor pixel value is higher than the central pixel value then value 1 is allocated to that position otherwise value 0 is allocated. The values are then read either clockwise or counter-clockwise into

a binary value. The aim of Thresholding is to get the local binary differences [88] that results in an eight-bit binary number.

The LBP operator is computationally simple and can withstand monotonic gray-scale changes[89]. Figure 2.6 shows a sample 3*3 input image and its corresponding LBP code. The generated LBP code is then used to obtain a global histogram.

Example

7	6	5
7	7	4
7	9	8

threshold

1	0	0
1		0
1	1	1

Weights

1	2	4
128		8
64	32	16

Pattern = 11110001_2

Decimal= $1+16+32+64+128=241_{10}$

Figure 2. 1 : Sample LBP code

The ability to discriminate and computational simplicity of LBP has made it a common technique for breast segmentation and classification. LBP has been adopted for identifying breast abnormalities in mammograms [90] which has facilitated LBP feature extraction during segmentation and classification of breast cancer tumors.

E. Local Ternary Patterns

Relying on the central pixel as a threshold in LBP makes it sensitive to noise. A minor change of central pixel significantly changes the LBP code. To overcome this challenge, LTP extends LBP by thresholding the pixels into (0, 1,-1) instead of (0, 1). The use of three value pixels makes LTP robust to noise than LBP. Consider threshold constant c, center pixel r, and neighbor pixel n. The LTP formula is given by:

$$S(N) = \begin{cases} 1, if\ n > r + c \\ 0, if\ n > r - c\ and\ r + c \\ -1\ if\ n < r - c \end{cases} \quad (3)$$

Where $S\ (N)$ s is the n^{th}neighbor containing the LTP code value.

After thresholding, to get rid of the negative values, the ternary pattern generated is divided into positive and negative patterns.

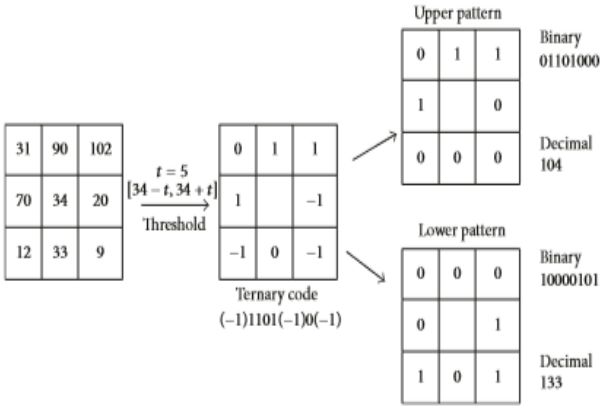

Figure 2. 2 : LTP Operator

While LTP has been able to resist noise, its major disadvantage is that modification done is invariant under grayscale transform of intensity values. Furthermore, the discriminate property is still inadequate since two or more different patterns can be falsely classified.

F. Local Directional Patterns

To resolve the challenge of relying on neighboring pixel intensity which makes LBP unstable, Jabid et al [91] proposed a Local Directional Pattern descriptor that encodes image texture by computing edge response values of a pixel in a different direction. Popular edge detectors like Frei-Chen, Kirsch, Sobel and Prewitt edge detectors are used in this case[92]. Among the edge detectors, Kirsch is the most popular because it identifies different directional edge responses more accurately by considering all the eight neighbors of a pixel[92]. Figure 2.8 shows the Kirsch mask.

$$\begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix} \begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix} \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} \begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}$$

$$M_0(0^\circ) \qquad M_1(45^\circ) \qquad M_2(90^\circ) \qquad M_3(135^\circ)$$

$$\begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix} \begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix} \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix} \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix}$$

$$M_4(180^\circ) \qquad M_5(225^\circ) \qquad M_6(270^\circ) \qquad M_7(315^\circ)$$

Figure 2. 3 : Kirsch Mask

The original LDP does not consider all response values because they are not all equally important. Therefore the k most top prominent directional values b_i are selected while the remaining $(8-k)$ are set to 0 as shown by the formula below.

$$LDP_k \sum_{i=0}^{7} bi\ (mi - mk) x\ 2^i \quad (1)$$

$$bi(a) = \begin{cases} 1\ a \geq 0 \\ 0\ a < 0 \end{cases} \quad (2)$$

85	32	26	313	97	503	0	0	1
53	50	10 →	537	X	399 →	1	X	1
60	38	45	161	97	303	0	0	0

LDP binary code: 00010011 LDP decimal code:19

Figure 2. 4 : A sample LDP code using $k = 3$

LDP operator is more robust to changes that may be as a result of noise due to non-monoatomic variations in illumination than LBP because LDP applies orientation responses that are more stable than intensity values used by LBP. Furthermore, LDP in the presence of noise produces the same pattern values[91]. Even though LDP is more stable than LBP, its reliance top three responses results in information loss around a local neighborhood. Also, according to [93] LDP is computationally expensive when compared to other techniques.

G. Local Quinary Patterns

While LBP and LTP encodes the pixels into (0,1) and (0,1,-1) respectively, Local Quinary Pattern (LQP) encodes the pixels into 5 values by using two threshold values τ_1 and τ_2 set by the user[94]. The formula is as shown

$$\begin{cases} 2 & p \geq gc + \tau_2 \\ 1 & gc + \tau_1 \leq p \leq gc + \tau_2 \\ 0 & gc - \tau_1 \leq p \leq gc + \tau_1 \\ -1 & gc - \tau_2 \leq p \leq gc - \tau_1 \\ -2 & Otherwise \end{cases} \quad (1)$$

The Local Quinary pattern is split into four binary patterns. Figure 2.10 illustrates how a Quinary code is split into four binary patterns.

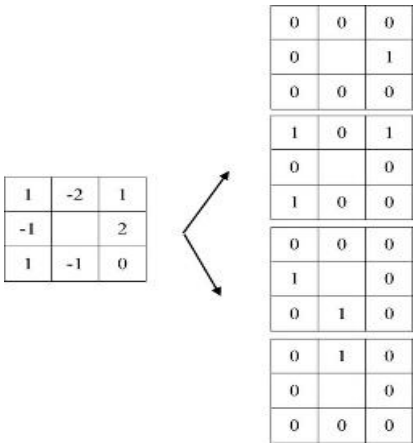

Figure 2. 5 : Quinary Code split into four binary code

VII. BREAST CANCER CLASSIFICATION TECHNIQUES

Classification is a vital stage in a CAD system that is meant to differentiate and label abnormalities by mapping data to a predefined target class. The objective of a classification algorithm is intended to build a classifier that takes some input features during training and learns the pattern then uses the knowledge learned to predict some new features during the testing phase [95]. Breast cancer is purely a classification problem in which the classifier assigns a label based on the attributes evaluated. It can either be a binary classification problem where two distinctive features are used to distinguish absences or presences of cancer cells or a multi-class problem in which more than two labels are used. Since radiologists are often prone to making mistakes when analyzing mammograms, using classification algorithms is a suitable approach for automating analysis of breast tumors because classifiers can learn complex relationships and patterns [96].

Classification techniques used in breast cancer detection can either be supervised or unsupervised. In supervised classification techniques, the classifier learns based on a training set and the knowledge gained is used for classification. Geometric and statistical distance measures are used to define how close a point is to each of the training samples. The approach has two main phases: A training phase in which data is analyzed by a classifier and a testing phase in which a classifier estimates accuracy using test data [97]. Supervised techniques rely on training set which distinguishes spectral distinctiveness of the classes by avoiding two or more classes being similar so that the rate of misclassification is low. To achieve this, a well-developed training set is required to ensure a good representation of data which aids good performance for the classifier. Some examples of commonly used supervised classification techniques for breast cancer classification include Support Vector Machine, Artificial Neural Networks, K-Nearest Neighbor, Random Forest, Bayesian and Naïve Bayes.

In unsupervised classification, there are no class samples provided, rather clusters are generated then the clusters are assigned classes (classified). It is faster than a supervised technique because prior learning is not required. Also, the classes are created based on spectral information only and therefore, they are not subjective. However, sometimes the spectral classes may not correspond with the information classes, consequently, more time is used to interpret and label the classes.

VIII. MAMMOGRAPHIC BREAST CANCER DATABASES

The purpose of breast cancer databases is to be able to facilitate research in breast cancer analysis and aid in developing algorithms that can be used in teaching and training. Currently, there are several breast cancer databases available for research. While some are publically available others are privately own and can only be accessed by specific institutions. Even though the research community has attempted to define an ideal breast cancer database, it is still evident that each database is unique in terms of the kind and number of cases handled, how images were taken and information provided about each case. Due to differences in strengths and weaknesses, it is a challenge to compare performance differences of models, approaches, and techniques based on these databases. However, a study by [1] indicates that DDSM and MIAS databases have the highest number of research papers published.

A. Digital Database for Screening Mammography (DDSM)

Digital Database for Screening Mammography is a publically available database [98] for researchers investigating mammogram image analysis. It was accessible through http://marathon.csee.usf.edu/Mammography/Database.html [99] which is no longer supported. The cases are organized into several volumes available online by anonymous ftp of ftp://figment.csee.usf.edu/pub/DDSM/cases.

The cases in the database were acquired from Massachusetts General Hospital, Wake Forest University School of Médicine, and Washington University. Originally, the database had 596 cases with 373 non-cancerous and 223 cancerous cases[99]. However, with time improvement has been made by adding more cases, currently, it has 2620 cases. Out of the 2620 cases, 695 are normal, 95 Benign and 101 Malignant. Also, an enhanced interface to enable previewing of each case has been developed. The file image type is Joint Pictures Expert Group (JPEG) and is 231GB in size distributed in 43 volumes. Every case consists of two images of each breast and patient information from different ethnic and racial backgrounds. The only information included is patient

age. However, a description of breast density using (ACR) and Breast Imaging Reporting and Data System (BI-RADS) annotation is also provide [100][101].

B. Mammographic Image Analysis Society (MIAS)

Mammographic Image Analysis Society is a 2.3 GB dataset generated by UK research groups that wanted to understand mammograms. It is publically accessible and contains 322 mammographic images of both breasts from 161 patients. The 322 digitized mammograms are in the mediolateral view out of which 51 are in malignant class, 64 are in benign class and 207 in the normal class [102]. Further, the MIAS database gives information on the location and radius of the tumor, type of abnormality and breast tissue affected [49]. The image file format is Portable Gray Map (PGM) and the mammograms are available through a Pilot European Image Processing (PEIPA) archives at the University of Essex.

Even though the database contained breast density information, Rungayyan [103] observed that breast density information was not classified according to any standard. Also, although MIAS annotation is defined based on the region of interest centered around a circular radius this is not enough for studies like [104] in which all circumscribed and spiculated masses had to be manually segmented. Furthermore, the resolution used to digitize the image makes MIAS unsuitable for experiments for detecting microcalcifications yet the database had the highest percentage of microcalcification [101].

C. Image Retrieval in Medical Applications (IRMA)

Images in Image Retrieval in Medical Applications (IRMA) database were collected from the Department of Diagnostic Radiology, Department of Medical Informatics, Division of Medical Image Processing and the Chair of Computer Science VI at the Aachen University of Technology [105]. The aim of the project was to develop and implement methods for Content-Based Image Retrieval. The database contains 10,509 reference images split into normal cases (12 volumes), cancer cases (15 volumes) and benign cases (14 volumes): each case may have one or more associated Pathological Lesions (PLs) segmentations, usually in Medio-lateral-Oblique (MLO) and Cranio-caudal (CC) images of the same breast [2].

D. Breast Cancer Digital Repository (BCDR)

Breast Cancer Digital Repository is a Portuguese breast cancer image database with real female patient information from medical records supplied by the Faculty of Medicine at the University of Porto, Portugal[106][98]. It was released for the public domain in 2012. However, it is still under development It has a total of 1734 patient cases against a total of 5776 Mediolateral-Oblique (MLO) and Cranio-Caudal (CC) image views[106].

The BCDR split into BCDR-FM and BCDR –DM BCDR-FM handles film mammography which constitutes 1010 (998 Female and 1Male) cases of the total 1734 with age 20 -90 years[106]. BCDR –DM is for Digital mammography and it is composed of 724 cases of which 723 are female and 1 male with ages between 27 and 92 years old.

E. INBreast

It was acquired at the breast center in CHSJ, Porto, authorized by Hospital Ethics Committee and National Committee of Data protection. It is a publicly available database with 115 cases and 410 MLO and CC images from which 90 cases are from women with both breasts affected (four images per case) and 25 cases from mastectomy patients (two images per case)[101]. The images were saved in DICOM format with all confidential information removed. The images are FFDM taken from screening, diagnostics and follow up cases. The database is available at

http://medicalresearch.inseporto.pt/breastresearch/GetINbreastDatabase.html.

F. Nijmegen database

The Nijmegen database contains 40 digitized mammograms of both CranioCaudal (CC) and Medio-Lateral Oblique (MLO) views recorded from 21 patients from the National Expert and Training Centre for Breast Cancer Screening and the Department of Radiology at the University of Nijmegen, the Netherlands. [107]. The images are digitized with a screen-film resolution of 2048 2048 pixels[107]. Each mammogram shows one or more microcalcification clusters verified by histology. The total number of clusters in the database is 105. The location and size of each cluster is given by a truth circle marked by expert radiologists. The database also contains look-up tables for rescaling of the image data and text files, storing the center and diameter of the truth circles in pixel coordinates[107].

G. Mammography Image reading for Radiologists and Computers Learning (MIRAcle) database

It is a web-accessible mammographic database that is made of a dynamic repository for machines and radiologists training and evaluation module[101]. It has 204 mammograms collected from 196 patients. The database is available for classification or education evaluation by the radiologist [108].

Other mammographic databases such as Lawrence Livermore National Laboratories (LLNL) [105] [109], Washington University Digital Mammography database [109], Trueta [110][111], Malaga[111] and Rheinisch-Westfälische Technische Hochschule (RWTH)[105] databases are found in literature. However most of them are not available, others are privately owned while others are still at the experimental stage therefore much information about them is known. More explanation about the can be found in [51, 59-61].

Table1.1: Features of breast cancer mammographic databases

Database	No. of images	Mode of image acquisition	Image view	Lesion type	No. of patients	Access type	Ground Truth
DDSM	9916	Screen Film	MLO and CC	All kinds	2620	Public	Boundary chain code of findings
MIAS	322	Screen Film	MLO	All kinds	161	Public	Centre & radius of circle around the area of interest
IRMA	10509	Screen Film	MLO and CC	All kinds	Unknown	Public	Boundary chain code of findings
BCDR	5776	FFDM	MLO, CC, and Ultrasound	All kinds	1734	Public	Lesion Contour
INBreast	410	FFDM	MLO and CC	All kinds	115	Public	Lesion Contour
Nijimegen	40	Screen Film	MLO and CC	MCs	21	Private	Centre & radius of circle around the area of interest

Banco Web LAPIMO	1400	Screen Film	MLO, CC, and others	All kinds	320	Public	ROI available in some images
MIRAcle	204	Unknown	Unknown	Unknown	196		ROI of findings
LLNL	198	Unknown	MLO and CC	MCs	50	Public	Binary images of MCs clusters &area of some MCs
Trueta	320	FFDM	Unknown	Unknown	89	Private	Centre & radius of circle around the area of interest
Malaga	Unknown	Unknown	MLO and CC	Masses	35	Unknown	Not available

IX. REVIEW OF MAMMOGRAPHIC BREAST CANCER DETECTION STUDIES

Shape, texture and hybrid features have been previously used for breast cancer detection. Using shape features to detect microcalcification, Zyout [55] used the SVM classifier with combined images from Mini MIAS and BMH databases. He achieved a classification accuracy of 100% using Leave-One-Out cross-validation. To detect microcalcification Boulehmi and Hourami [46] proposed a technique for microcalcification detection based on generalized Gaussian density. Using three shape descriptors and a neural fuzzy system to distinguish between benign /malignant lesions, they achieved an accuracy of 94.44%. Surendiran and Vadivel [112] used various geometric features using DDSM with 1553 images with a CART classifier. They obtained 93.72% accuracy.

Using texture features (Gabor filter and LBP), Sansare and Kinge [89] classified benign, malignant and normal breast cells using SVM classifier on 158 MIAS images. They achieved accuracy of 89.28%, 70.37% and 79.61% for the benign, malignant and normal class respectively. To detect microcalcification Arai et al [113] employed multi-branch standard deviation analysis and surrounding region dependence method. They achieved an overall accuracy of 70.8% using the Japanese society of medical imaging technology database with Neural Networks. Krishnaveni [114] achieved a higher accuracy of 96.25% than Arai et al[113] by using Naïve Bayes and MIAS databases. Instead of using the entire breast region, Rampun et al [115] employed an LTP descriptor that modeled the appearance of the fibro-granular disk region. They used SVM on the MIAS database with stratified ten-fold cross-validation. They achieved an accuracy of 82.33%. Herwanto and Arymurthy [72] applied GLCM features to model a mass breast cancer classification system for normal/abnormal tumors using 73 images from the MIAS database. They achieved an accuracy of 86% and 90% for the normal/abnormal classes respectively.

Combining shape and texture features provides good results and as such Alharbi et al [116] developed a feature reduction framework by extracting 49 out of 109 intensity, shape and texture features. Using the IRMA database with Feed Forward Neural Network, they achieved an accuracy of 94.27%. Using the KNN classifier with DDSM images, Sudha and Selvarajan [117]

proposed a breast cancer classification system using a minimal number of features from shape texture and intensity. They achieved a classification accuracy of 99.13%. Oliver et al [118] used fuzzy C mean clustering to segment the breast region and extract shape and texture features from each cluster. Sequential Forward selection with KNN was used. They achieved an accuracy of 86%. Parthian et al [119] used a similar approach but with a more sophisticated feature selection approach and achieved an accuracy of 91.4%.

X. CONCLUSION

A comprehensive survey of several techniques and approaches used for breast cancer detection has been presented. A lot of effort has been dedicated to the academic literature on, development and classification of breast cancer, mammographic technologies, mammographic signs of breast cancer, breast cancer datasets, feature extraction and classification techniques.

The major indicators of breast cancer tumor is through mass and microcalcification detection. Though mass detection seems to be more challenging than microcalcification because of poor image contrast, a majority of studies opted to use mass detection which requires image enhancement to improve image contrast. The choice of using mass over microcalcification could be a attributed to high accuracy levels achieved by previous studies.

In breast cancer modeling approaches, hybrid modeling achieves better results than appearance or shape modeling. However, care should be taken to ensure no two or more similar models are combined as no much improvement will be achieved. Also while modeling a hybrid model care should be taken such that no two or more combined models cancel the effects of each other. Models are combined either in parallel, in which individual outputs of the models are concatenated and used for final breast cancer detection or hierarchical in which the output of one model becomes the input of the other modeling technique.

Extracting discriminative features from a mammogram image is extremely important because if the features are not discriminative enough there is a high rate of false positives. Existing breast cancer feature extraction approaches are based on shape, texture or a combination of both. Local texture feature extraction approaches have not been widely applied to breast cancer detection as in face recognition. This could be attributed to the fact that this feature extraction approaches produce low accuracy levels as seen in [74],[120] and those that attempted to produce high accuracy levels are too complex. Also, combining two or more techniques is expected to yield better results, however in some circumstances, this combination has produced the worst results than individual technique.

To solve the problem of breast cancer classification a variety of classifiers have been applied in different ways. The most employed classifier being SVM, ANN, and KNN. Popularity of SVM could be attributed to the ability to extend the patterns that are not linearly separable by using a kernel function which transforms a low dimensional input space to a higher-dimensional space. Also, its robustness to outliers because it ignores outliers and maximizes the margin around the separating hyperplane. On the other hand, while KNN may not handle outliers well like SVM itis a simple non-parametric method that can detect linear and nonlinear distributed data and perform very well with a lot of data points. In comparison to SVM, ANN suffers from multiple local minima, however, ANN can model complex relationships and generalize and make a prediction on unseen data.

Several publically and privately accessible databases with different numbers of images exist for breast cancer classification. It was observed that the most easily accessed databases and therefore most commonly used databases are the Mammographic Image Analysis Society (MIAS) database and Digital Database for Screening Mammography (DDSM). Also, for some of the studies in which classification accuracy level was rated highly could have been as a result of using databases with few images or working with fewer images within a chosen database.

Future work that may improve breast cancer performance is Fusion. Feature and decision fusion in breast cancer detection has not been fully exploited. Fusing shape, texture, intensity and dense features could result in a feature set with more discriminate features that can differentiate different tumor types. Further ensemble of classifiers can result in better decisions which leads to better classification accuracy.

XI. REFERENCES

[1] A. Nahid and Y. Kong, "Involvement of Machine Learning for Breast Cancer Image Classification : A Survey," *Hindawi Comput. Math. Methods Med.*, vol. 2017, no. i, p. 29, 2018.

[2] N. P. Pérez, "Improving Variable Selection and Mammography-based Machine Learning Classifiers for Breast Cancer CADx," 2015.

[3] P. Skaane, "Radiology Screen-Film Mammography versus Full-Field Digital Mammography with Soft-Copy Reading : Randomized Trial in a Population-based Screening Program — The Oslo II Study 1," pp. 197–204, 2004.

[4] N. M. Hambly, N. Phelan, and F. L. Flanagan, "Women's Imaging • Or iginal Research Comparison of Digital Mammography and Screen-Film Mammography in Breast Cancer Screening: A Review in the Irish Breast Screening Program," no. October, pp. 1010–1018, 2009.

[5] N. M. Hambly, M. M. McNicholas, N. Phelan, G. C. Hargaden, A. O'Doherty, and F. L. Flanagan, "Comparison of digital mammography and screen-film mammography in breast cancer screening: A review in the Irish Breast Screening Program," *Am. J. Roentgenol.*, vol. 193, no. 4, pp. 1010–1018, 2009.

[6] M. R. Del Turco *et al.*, "Full-field digital versus screen-film mammography: Comparative accuracy in concurrent screening cohorts," *Am. J. Roentgenol.*, vol. 189, no. 4, pp. 860–866, 2007.

[7] J. M. Lewin *et al.*, "Clinical comparison of full-field digital mammography and screen-film mammography for detection of breast cancer," *AJR Am J Roentgenol*, vol. 179, no. 3, pp. 671–677, 2002.

[8] J. Ferlay *et al.*, "Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012," *Int. J. Cancer*, vol. 136, no. 5, pp. E359–E386, 2015.

[9] F. Bray, J. Ferlay, and I. Soerjomataram, "Global Cancer Statistics 2018 : GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA Cancer*, pp. 394–424, 2018.

[10] "Breast Cancer," 2018.

[11] A. Korir, N. Okerosi, V. Ronoh, G. Mutuma, and M. Parkin, "Incidence of cancer in Nairobi, Kenya (2004-2008)," *Int. J. Cancer*, vol. 137, no. 9, pp. 2053–2059, 2015.

[12] M. A. Berbar, "Hybrid methods for feature extraction for breast masses classification," *Egypt. Informatics J.*, vol. 19, no. 1, pp. 63–73, 2018.

[13] A. Rampun, "Breast Density Classification Using Local Quinary Patterns with Various Neighbourhood Topologies †," *J. imaging*, pp. 1–23, 2018.

[14] K. U. Sheba and S. G. Raj, "An approach for automatic lesion detection in mammograms," *Cogent Eng.*, vol. 2016, pp. 1–16, 2018.

[15] B. Singh and M. Kaur, "An approach for classification of malignant and benign microcalcification clusters," *Sādhanā*, vol. 43, no. 3, pp. 1–18, 2018.

[16] Breast cancer UK, "Biology of Breast Cancer," no. 2, 2016.

[17] J. D. Davis and S. Lin, "DNA damage and breast Cancer," *World J. Clin. Oncol.*, vol. 2, no. 9, pp. 329–338, 2011.

[18] I. Jatoi and M. Kaufmann, *Management of Breast Diseases*. 2010.

[19] G. Sharma, R. Dave, jyotsana Sanadya, S. Piush, and K. . Sharma, "Various types and management of breast cancer: An Overview," *J. Adv. Phamaceutical Technol. Res.*

[20] R. Reilly, "Breast Cancer," no. 1996, pp. 1–9, 2007.

[21] M. Vasantha, "Classifications of Mammogram Images Using Hybrid Features," 2015.

[22] M. N. Patel and P. Tandel, "A Survey on Feature Extraction Techniques for Shape based Object Recognition," *Int. J. Comput. Appl.*, vol. 137, no. 6, pp. 16–20, 2016.

[23] E. Omidiora, S. O. Olabiyisi, A. Temitope, and A. Temilola, "Feature Extraction Techniques for Mass Detection in Digital Mammogram (Review)," *J. Sci. Res. Reports*, no. February 2019, 2017.

[24] S. Caulkin and S. Astley, "Generating Realistic Mass Lesions In Digital Mammograms Using Statistical models," pp. 285–294, 1999.

[25] M. Berks, S. Caulkin, R. Rahim, C. Boggis, and S. Astley, "Statistical Appearance Models of Mammographic Masses," pp. 401–408.

[26] X. Zhang, J. Cui, W. Wang, and C. Lin, "A study for texture feature extraction of high-resolution satellite images based on a direction measure and gray level co-occurrence matrix fusion algorithm," *Sensors (Switzerland)*, vol. 17, no. 7, 2017.

[27] P. Delogu, M. Evelina, P. Kasae, and A. Retico, "Characterization of mammographic masses using a gradient-based segmentation algorithm and a neural classifier," vol. 37, pp. 1479–1491, 2007.

[28] M. P. Sampat, M. K. Markey, and A. C. Bovik, *Computer-Aided Detection and Diagnosis in Mammography*, Second Edi. Elsevier Inc.

[29] M. Varma and A. Zisserman, "A Statistical Approach to Texture Classification from Single Images," 2004.

[30] J. Tang, S. Member, R. M. Rangayyan, J. Xu, and I. El Naqa, "Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography : Recent Advances," *IEEE Trans. onInformation Technol. Biomed.*, vol. 13, no. 2, pp. 236–251, 2009.

[31] H. M. H. Alharbi, P. Kwan, A. Jayawardena, and A. S. M. Sajeev, "Fuzzy Image Segmentation for Mass Detection in Digital Mammography," *Multidiscip. Comput. Intell. Tech.*, no. January, pp. 378–402, 2012.

[32] N. Köşüş, A. Köşüş, M. Duran, S. Simavli, and N. Turhan, "Meme kanseri taramasi{dotless}nda standart mamografi ile dijital mamografi ve dijital infrared termal görüntülemenin karşi{dotless}laşti{dotless}nlmasi{dotless}," *J. Turkish Ger. Gynecol. Assoc.*, vol. 11, no. 3, pp. 152–157, 2010.

[33] G. I. Andreea, R. Pegza, L. Lascu, S. Bondari, Z. Stoica, and A. Bondari, "The Role of Imaging Techniques in Diagnosis of Breast Cancer," *Curr. Heal. Sci. J.*, vol. 37, no. 2, pp. 55–61, 2011.

[34] S.-T. Luo and B.-W. Cheng, "Diagnosing Breast Masses in Digital Mammography Using Feature Selection and Ensemble Methods," *J. Med. Syst.*, vol. 36, no. 2, pp.

569–577, 2012.

[35] E. D. Pisano *et al.*, "new england journal," *N. Engl. J. Med.*, pp. 1773–1783, 2005.

[36] P. Skaane, "Randomized Trial of Screen-Film versus Full-Field Digital Reading in Population-based Screening Program : Follow-up and Purpose : Methods : Results : Conclusion :," vol. 244, no. 3, 2007.

[37] I. Juel, P. Skaane, S. R. Hoff, and G. Johannessen, "Screen-film mammography versus full-field digital mammography in a population-based screening program : The Sogn and Fjordane study Screen-fi lm mammography versus full-fi eld digital mammography in a population-based screening program : The Sogn and Fjor," vol. 1851, 2010.

[38] E. D. Pisano *et al.*, "Diagnostic Accuracy of Digital versus Film Mammography: Exploratory Analysis of Selected Population Subgroups in DMIST," *Radiology*, vol. 246, no. 2, pp. 376–383, 2008.

[39] A. Jalalian *et al.*, "Review article : Foundation and Methodologies in Computer Aided diagnosis systems for breast cancer detection," pp. 113–137, 2017.

[40] B. C. Maria Rizzi, Matteo D'Aloia, "Review: Health care CAD system for breast microcalcification cluster detection," *J. Med. Biol. Eng.*, vol. 32, no. 3, pp. 147–156, 2011.

[41] M. Posso and T. Puig, "Cost-Effectiveness of Double Reading versus Single Reading of Mammograms in a Breast Cancer Screening Programme," 2016.

[42] M. C. Posso, T. Puig, M. J. Quintana, J. Solà-roca, and X. Bonfill, "Double versus single reading of mammograms in a breast cancer screening programme : a cost-consequence analysis," 2016.

[43] P. Chu, K. Kerlikowske, E. A. Sickles, and R. Smith-bindman, "Positive Predictive Value of Specific Mammographic Findings according to Reader Methods : Results : Conclusion :," vol. 250, no. 3, 2009.

[44] C. Gajdos *et al.*, "Mammographic Appearance of Nonpalpable Breast Cancer Reflects Pathologic Characteristics," *Ann. Surg.*, vol. 235, no. 2, pp. 246–251, 2002.

[45] J. Bozek, M. Mustra, K. Delac, and M. Grgic, "A Survey of Image Processing Algorithms in Digital Mammography," pp. 631–657, 2009.

[46] H. Boulehmi, H. Mahersia, and K. Hamrouni, "A New CAD System for Breast Microcalcifications Diagnosis," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 4, pp. 133–143, 2016.

[47] I. Zyout, I. Abdel-qader, and C. Jacobs, "Bayesian Classifier with Simplified Learning Phase for Detecting Microcalcifications in Digital Mammograms," *Int. J. Biomed. Imaging*, vol. 2009, 2009.

[48] S. O. Grady and M. P. Morgan, "BBA - Reviews on Cancer Microcalci fi cations in breast cancer : From pathophysiology to diagnosis and prognosis," vol. 1869, no. March, pp. 310–320, 2018.

[49] M. a. Alolfe, W. a. Mohamed, A.-B. M. Youssef, Y. M. Kadah, and A. S. Mohamed, "Feature selection in computer aided diagnostic system for microcalcification detection in digital mammograms," in *26th National Radio Science Conference (NRSC2009)*, 2009.

[50] T. Balakumaran, "Detection of Microcalcification in Mammograms Using Wavelet Transform and Fuzzy Shell Clustering," vol. 7, no. 1, pp. 121–125, 2010.

[51] M. G. Mini, V. P. Devassia, and T. Thomas, "Multiplexed Wavelet Transform Technique for Detection of Microcalcification in Digitized Mammograms," *J. Digit. Imaging*, vol. 17, no. 4, pp. 285–291, 2004.

[52] D. Gunawan, "Microcalcification Detection Using Wavelet Transform," pp. 694–697.

[53] S. Bouyahia, J. Mbainaibeye, and N. Ellouze,

"Wavelet Based Microcalcifications Detection in Digitized Mammograms," no. January, pp. 23–31, 2009.

[54] N. B. Karayiannis, "Detection of Microcalcifications in Digital Mammograms Using Wavelets," no. September, 2014.

[55] I. Zyout, "Computer-Aided Diagnosis of Microcalcification Clusters Using Morphology Based Features and PSO-SVM Parameter Selection," vol. 2, no. 2, pp. 126–144, 2016.

[56] I. Zyout, I. Abdel-qader, and C. Jacobs, "Embedded Feature Selection using PSO-kNN: Shape-Based Diagnosis of Microcalcification Clusters in Mammography.," *Juspn*, vol. 3, no. 1, pp. 7–11, 2011.

[57] P.-P. T. 1 Yi-Jhe Huang 1, Ding-Yuan Chan 2, Da-Chuan Cheng 3,*, Yung-Jen Ho 3,*, "Automated Feature Set Selection and Its Application to MCC Identification in Digital Mammograms for Breast Cancer Detection," pp. 4855–4875, 2013.

[58] K. Geethal and K. T. A. Kishore, "New Particle Swarm Optimization for Feature Selection and Classification of Microcalcifications in Mammograms Method Anlyl," *IEEE-International Conf. Signal Process. Netw. Madras Inst. Technol. Anna Univ. Chennai India,* vol. 4, no. 6, pp. 458–463, 2008.

[59] M. P. Sampat, M. K. Markey, and A. C. Bovik, "Computer-Aided Detection and Diagnosis in Mammography," *Handb. Image Video Process.*, pp. 1195–1217, 2005.

[60] S. Liu, C. F. Babbs, and E. J. Delp, "Multiresolution detection of spiculated lesions in digital mammograms," *IEEE Trans. Image Process.*, vol. 10, no. 6, pp. 874–884, 2001.

[61] M. P. Sampat and A. C. Bovik, "Detection of Spiculated Lesions in Mammograms," *Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc.*, vol. 1, no. July, pp. 810–813, 2003.

[62] R. Zwiggelaar *et al.*, "Model-based detection of spiculated lesions in mammograms," *Med. Image Anal.*, vol. 3, no. 1, pp. 39–62, 1999.

[63] J. Wei *et al.*, "Computer-aided detection of breast masses on mammograms: Dual system approach with two-view analysis," *Med. Phys.*, vol. 36, no. 10, pp. 4451–4460, 2009.

[64] A. Elmoufid *et al.*, "Automatic Diagnosing of Suspicious Lesions in Digital Mammograms," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 5, pp. 510–518, 2016.

[65] S. Punitha, A. Amuthan, and K. S. Joseph, "Benign and malignant breast cancer segmentation using optimized region growing technique," *Futur. Comput. Informatics J.*, vol. 3, no. 2, pp. 348–358, 2018.

[66] K. Hu, X. Gao, and F. Li, "Detection of suspicious lesions by adaptive thresholding based on multiresolution analysis in mammograms," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 2, pp. 462–472, 2011.

[67] F. Soares Sérvulo de Oliveira, A. Oseas de Carvalho Filho, A. Corrêa Silva, A. Cardoso de Paiva, and M. Gattass, "Classification of breast regions as mass and non-mass based on digital mammograms using taxonomic indexes and SVM," *Comput. Biol. Med.*, vol. 57, pp. 42–53, 2014.

[68] G. M. Te Brake, N. Karssemeijer, and J. H. C. L. Hendriks, "An automatic method to discriminate malignant masses from normal tissue in digital mammograms," *Phys. Med. Biol.*, vol. 45, no. 10, pp. 2843–2857, 2000.

[69] Y. Wang, F. Aghaei, A. Zarafshani, Y. Qiu, W. Qian, and B. Zheng, "Computer-aided classification of mammographic masses using visually sensitive image features," *J. Xray. Sci. Technol.*, vol. 25, no. 1, pp. 171–186, 2017.

[70] N. R. Mudigonda, R. M. Rangayyan, and J.

E. L. Desautels, "Detection of Breast Masses in Mammograms by Density Slicing and Texture Flow-Field Analysis," vol. 20, no. 12, pp. 1215–1227, 2001.

[71] R. Nithya and B. Santhi, "Mammogram classification using maximum difference feature selection method," *J. Theor. Appl. Inf. Technol.*, vol. 33, no. 2, 2011.

[72] Herwanto and A. Arymurthy, "A System for Computer Aided Diagnosis of Breast Cancer Based on Mass Analysis," *Int. Conf. Robot. Biomimetics, Intell. Comput. Syst. Yogyakarta,* pp. 247–253, 2013.

[73] X. Liu, X. Xu, and J. Liu, "A new automatic method for mass detection in mammography with false positives reduction by supported vector machine," pp. 33–37, 2011.

[74] H. A. Khan, A. Al Helal, K. I. Ahmed, and R. Mostafa, "Abnormal Mass Classification in Breast Mammography using Rotation Invariant LBP," no. September, 2016.

[75] R. Nithya and B. Santhi, "Computer Aided Diagnosis System for Mammogram Analysis : A Survey," vol. 5, no. 4, 2015.

[76] R. J. McKenna, "The abnormal mammogram radiographic findings, diagnostic options, pathology, and stage of cancer diagnosis," *Cancer*, vol. 74, no. 1 S, pp. 244–255, 1994.

[77] S. Banik, R. M. Rangayyan, and J. E. Leo Desautels, "Computer-aided detection of architectural distortion in prior mammograms of interval cancer," *J. Digit. Imaging*, vol. 47, no. 5, pp. 1–193, 2013.

[78] M. K. Shetty, "Mammographic Signs of Breast Cancer," pp. 93–117, 2015.

[79] V. Lattanzio and G. Simonetti, *Mammography: Guide to Interpreting, Reporting and Auditing Mammographic Images*, vol. 240, no. 2. 2006.

[80] R. Bhanumathi and G. . Suresh, "Performance Analysis in Computer Aided Detection of Breast Cancer by Mammography," *Int. J. IT Eng.*, vol. 01, no. 02, 2013.

[81] M. J. M. Broeders, N. C. Onland-Moret, H. J. T. M. Rijken, J. H. C. L. Hendriks, A. L. M. Verbeek, and R. Holland, "Use of previous screening mammograms to identify features indicating cases that would have a possible gain in prognosis following earlier detection," *Eur. J. Cancer*, vol. 39, no. 12, pp. 1770–1775, 2003.

[82] R. M. Rangayyan, F. J. Ayres, and J. E. Leo Desautels, "A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs," *J. Franklin Inst.*, vol. 344, no. 3–4, pp. 312–348, 2007.

[83] Y. Zheng, "Breast Cancer Detection with Gabor Features from Digital Mammograms," *Open Access*, pp. 44–62, 2010.

[84] S. Khan and M. Hussain, "A comparison of different Gabor feature extraction approaches for mass classification in mammography," 2015.

[85] E. K. Sharma, E. Priyanka, E. A. Kalsh, and E. K. Saini, "GLCM and its Features," *Int. J. Adv. Res. Electron. Commun. Eng. Vol.*, vol. 4, no. 8, pp. 2180–2182, 2015.

[86] P. Mohanaiah, P. Sathyanarayana, and L. Gurukumar, "Image Texture Feature Extraction Using GLCM Approach," vol. 3, no. 5, pp. 1–5, 2013.

[87] T. Ojala, M. Pietikainen, and D. Harwood, "A Comparative Study on Texture Measures with Classification based on Feature Distributions," *Pattern Recognit.*, vol. 29, no. l, 1996.

[88] T. H. Rassem and B. E. Khoo, "Completed Local Ternary Pattern for Rotation Invariant Texture Classification," *Sci. World J.*, vol. 2014, 2014.

[89] A. V Sansare and S. R. Kinge, "Classification of Breast cancer Using Local Binary Pattern and Gabor Filter," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 5, no. X, pp. 1389–1395, 2017.

[90] S. Naresh and M. Tech, "Breast Cancer

Detection using Local Binary Patterns," vol. 123, no. 16, pp. 6–9, 2015.

[91] T. Jabid, H. Kabir, and O. Chae, "Local Directional Pattern (LDP) – A Robust Image Descriptor for Object Recognition," *Seventh IEEE Int. Conf. Adv. Video Signal Based Surveill.*, pp. 482–487, 2010.

[92] S. Lee, "Multilayer Cluster Neural Network for Totally Unconstrained Handwritten Numeral Recognition," vol. 8, no. 5, pp. 783–792, 1995.

[93] A. M. Shabat and J. Tapamo, "A comparative study of the use of local directional pattern for texture-based informal settlement classification," *J. Appl. Res. Technol.*, vol. 15, no. 3, pp. 250–258, 2017.

[94] L. Nanni, A. Lumini, and S. Brahnam, "Artificial Intelligence in Medicine Local binary patterns variants as texture descriptors for medical image analysis," *Artif. Intell. Med.*, vol. 49, no. 2, pp. 117–125, 2010.

[95] H. Karim and K. Zand, "A Comparative Survey on data mining techniques for breast cancer diagnosis and prediction," *Int. J. Fundam. Appl. Life Sci.*, vol. 5, no. 2005, pp. 4330–4339, 2015.

[96] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Supervised Machine Learning: A Review of Classification Techniques," *Informatica*, vol. 31, pp. 501–520, 2007.

[97] J. Han and M. Kamber, *Data mining : Concepts and Techniques*. 2006.

[98] D. C. Moura and M. A. Guevara, "An evaluation of image descriptors combined with clinical data for breast cancer diagnosis," pp. 561–574, 2013.

[99] K. Bowyer, D. Kopans, P. K. Jr, R. Moore, K. Chang, and S. Munishkumaran, "Current Status of the Digital Database for Screening Mammography," pp. 457–460.

[100] P. K. M. Heath, K. Bowyer, D. Kopans, R. Moore, "The Digital Database for Screening Mammography." pp. 1–10.

[101] C. Moreira, A. Cardoso, and S. Jaime, "INbreast:Toward a Full- Field Digital Mammographic Database," 2012.

[102] J. Sunkling, "Themammographic image analysis society digital mammogram database," no. January 1994. 2014.

[103] R. M. Rangayyan and N. R. I. Mudigonda, "Boundary modelling and shape analysis methods for classification of mammographic masses," *Med. Biol. Eng. Comput.*, vol. 38, 2000.

[104] A. Oliver, J. Freixenet, J. Martí, E. Pérez, J. Pont, and E. R. E. Denton, "A review of automatic mass detection and segmentation in mammographic images," *Med. Image Anal. Elsevier*, vol. 14, no. 2, pp. 87–110, 2010.

[105] J. E. E. Oliveira, M. O. Gueld, A. D. A. Araújo, B. Ott, and T. M. Deserno, "Towards a Standard Reference Database for Computer-aided Mammography," in *Proceedings of SPIE*, 2008, vol. 6915, pp. 1–9.

[106] M. Angel, G. Lopez, and D. C. Moura, "BCDR : A Breast Cancer Digital Repository," *Res. Gate*, no. September 2017, 2012.

[107] T. Netsch and H. Peitgen, "Scale-Space Signatures for the Detection of Clustered Microcalcifications in Digital Mammograms," vol. 18, no. 9, pp. 774–786, 1999.

[108] Z. C. Antoniou *et al.*, "A web- accesible mammographic image database dedicated to combine training and evaluation of radiologists and machines," 2009.

[109] J. asjit S. Suri, S. K. Setarehdan, and S. Singh, *Advanced Algorithms Approaches to Medical Image Segmentation*. 2002.

[110] A. Oliver, J. Pont, E. R. E. Denton, and J. Freixenet, "A Statistical Approach for Breast Density Segmentation," vol. 23, no. 5, pp. 527–537, 2010.

[111] A. O. Malagelada, *Automatic mass segmentation in mammographic images*, no. May. 2016.

[112] B. Surendiran and A. Vadivel, "Mammogram mass classification using various geometric shape and margin features for early detection of breast cancer," *Int. J. Med. Eng. Informatics*, vol. 4, no. 1, pp. 36–54, 2012.

[113] K. Arai, I. Nugraha, and H. Okumura, "Automated Detection Method for Clustered Microcalcification in Mammogram Image Based on Statistical Textural Features," *Int. J. Adv. Res. Artif. Intell.*, vol. 1, no. 3, pp. 12–16, 2012.

[114] S. Krishnaveni, R. Bhanumath, and T. Pugazharasan, "Study of Mammogram Microcalcification to aid tumour detection using Naive Bayes Classifier," *Int. J. Adv. Res. Electr. Electron. Instrum. Eng.*, vol. 3, no. 3, pp. 8274–8282, 2014.

[115] A. Rampun, P. Morrow, B. Scotney, and J. Winder, "Breast Density Classification Using Local Ternary Patterns in Mammograms," in *Research Gate*, 2017, no. June.

[116] H. Alharbi, G. Falzon, and P. Kwan, "A novel feature reduction framework for digital mammogram image classification," *2015 3rd IAPR Asian Conf. Pattern Recognit.*, pp. 221–225, 2015.

[117] M. N. Sudha and S. Selvarajan, "Feature Selection Based on Enhanced Cuckoo Search for Breast Cancer Classification in Mammogram Image," *Circuits Syst.*, vol. 7, no. 04, p. 327, 2016.

[118] A. Oliver, J. Freixenet, R. Mart, J. Pont, and P. Elsa, "A Novel Breast Tissue Classification Methodology," *IEEE Trans. Inf. Technol. Biomed.*, no. February, 2008.

[119] N. Mac Parthaláin, R. Jensen, Q. Shen, and R. Zwiggelaar, "Fuzzy-rough approaches for mammographic risk analysis," *Intell. Data Anal.*, vol. 14, no. 2, pp. 225–244, 2010.

[120] J. Liu, "Improved Local Binary Patterns for Classification of Masses Using Mammography," *IEEE*, pp. 2692–2695, 2011.

[121] Z. Lai and H. Deng, "Medical Image Classification Based on Deep Features Extracted by Deep Model and Statistic Feature Fusion with Multilayer Perceptron," *Comput. Intell. Neurosci.*, vol. 2018, 2018.

[122] J. M. Specht and D. A. Mankoff, "Advances in molecular imaging for breast cancer detection and characterization," *Breast Cancer Res.*, vol. 14, no. 2, 2012.

Cite this article as :

Mary Walowe Mwadulo, Raphael Angulu, Stephen Makau Mutua, "Breast Cancer Detection via Mammographic Images : A Survey ", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6, Issue 3, pp.173-195, May-June-2020. Available at
doi : https://doi.org/10.32628/CSEIT20633
Journal URL : http://ijsrcseit.com/CSEIT20633