Morphological and Textural Data Fusion for Breast Cancer Classification Based on Inter and Intra group Variances

Gurudas VR[1]   Shaila SG[1]*   Vadivel A[2]

[1]*Dayananda Sagar University, Bangalore, Karnataka, India*
[2]*GITAM School of Technology, GITAM University, Bangalore, Karnataka, India*
* Corresponding author's Email: shaila-cse@dsu.edu.in

Abstract: Nowadays, the most predominant cancer disease is Breast Cancer that has a higher death rate and women gender is the most affected by this disease. But detecting Breast Cancer in early stage is challenging as the malignance growth at this stage occurs in the duct that are undetected as symptoms are less. This paper addresses the challenge of early detection of Breast Cancer cells by proposing the fusion scheme of morphological and texture features of the cells for analysis. Morphological features such as the shape and marginal characteristics of the mass are considered as per the Breast Imaging Reporting and Data System (BI-RADS) standard. Texture features of the mass were also extracted to understand the characteristics of pixel variation in the masses. These features are combined and its dimension is normalized using Exhaustive Feature Selection (EFS). The accuracy of the proposed feature on the INbreast dataset is 94.75% on an average. The accuracy for the Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) Calc RoI dataset is 95% and for CBIS-DDSM Mass RoI dataset it is 94.5%. The result is further compared with contemporary methods and found that the fused feature is performing well.

Keywords: Breast cancer, Benign, Malignant, Mass, Shape, Texture, Fusion, CBIS-DDSM, INbreast.

1. Introduction

Breast Cancer is considered as the world's most predominant cancer disease with a higher death rate and it is known that women are the most affected gender by this disease. The Breast Cancer spreads all over the body and 1 among 8 women may be affected during their lifetime. World Health Organization (WHO) reports 2.3 million women were diagnosed in 2022 and there may be 6,85,000 deaths globally by the end of 2023 [1] (https://www.who.int/news-room/fact-sheets/detail/breast-cancer). Further, 7.8 million women have been diagnosed with Breast Cancer. In general, the malignance growth at the early stage occurs in the duct or lobule which are undetected as symptoms are less and has a low chance of spreading. Thus, early detection as well as treatment is challenging.

The most common and well-known screening techniques are Breast Self-Examination (BSE), Clinical Breast Examination (CBE) and Digital Mammography (DM) [2]. Digital Mammography is widely used as the primary screening tool for detecting Breast Cancer as it is least expensive and very sensitive to small lesions. In recent times, Computer-Aided Diagnosis (CAD) systems are used commonly to guide radiologists for Breast Cancer diagnosis. Earlier, CAD systems have relied on handmade visual information which had challenges in predicting Breast Cancer in those images. Recent growths of CAD systems have used the Artificial Intelligence techniques that automatically identifies breast tumors using Ultrasound images [3]. Computerized techniques involve various stages such as ultrasound images pre-processing followed by segmentation, feature extraction and classification.

This paper addresses the challenge of early detection of Breast Cancer cells by proposing the fusion scheme of the low-level features for analysis of cancer area. The proposed framework use Mammogram images for detecting and analyzing

breast cancer. The proposed framework uses morphological and texture features of the masses for analysis. The masses are segmented and localized using well-known scheme. Morphological features such as the shape and marginal characteristics of the mass are considered. However, these characteristics alone may not be sufficient to improve the classification accuracy. Thus, texture features of the mass were also extracted to analyse the pixel variation characteristics in the masses. Non-parametric test is performed on the pixel distribution and texture features are extracted through Goodness of Fit Distribution (GFD), Independence of Attributes (IoA) and Homogeneity (HG). These features are combined and its dimension is normalized using Exhaustive Feature Selection (EFS). The dimensionality reduction is done based on its contribution in discriminating the mass. The machine learning techniques are used for the classification and the performance of fused feature on Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) and INbreast datasets is found to be encouraging.

The rest of the paper is organized as follows. The literature is reviewed and presented in the next Section. The proposed approach is explained in Section 3, the experimental results are explained in Section 4 and the last Section concludes the paper.

2. Materials and methods

This section presents the reviews of the relevant and recent literatures on fusing various features. The authors in [4] have fused texture, morphological and histogram-based features for classifying tumors. However, the approach has taken more computational time it considered more features. The authors in [5] have extracted 41 morphological features and 96 texture features and classified the tumors using 7 classifiers. The performance resulted in poor accuracy. The authors in [6] have proposed a technique to segment breast masses based on the colour and intensity variation. The textural deviation and mathematical morphology have been obtained by extracting pixel features using colour histogram, however, the approach achieved in poor results. The authors in [7] have used segmentation methods to segment tumors in mammogram images. The segmentation methods have been grouped into classical segmentation, machine-learning based segmentation, supervised-unsupervised segmentation and deep-learning based segmentation. The classical segmentation includes edge-based, threshold-based and region-based. These segmentation schemes and the deep learning U-Net

model together have been used for training the mammogram images. The approach worked on small dataset. Authors in [8] have presented hybrid thresholding and used machine learning techniques to classify benign or cancerous cells in mammogram images. The approach has used hybrid thresholding to derive the RoI. The wavelet transform has been used to reduce the noise from each block based on Bayes Shrink soft thresholding. A Multi-Fractal Dimension (M-FD) has been applied to extract multiple features from each de-noised block. Genetic algorithm has been used to optimize the number of features. The approach took more computational time. The authors in [9] have used an Evolutionary Programming Neural Ensemble (EPNE) approach for the prediction of Breast Cancer using Fine-Needle Aspiration Biopsy (FNAB) samples. The predictive model considers various features of the FNAB samples, such as the size and shape of the cells to classify them into different categories of Breast Cancer.

The authors in [10] have proposed the DE-Ada* model for classifying the Breast Cancer. The authors have extracted a set of features to characterize mammograms and fused the features based on a weight mechanism. The approach resulted in poor precision and recall. The authors in [11] have presented a CAD method for detecting Breast Cancer. The authors used Hilbert Transform (HT) on rough data for reconstructing brightness-mode images. Marker-controlled watershed transformation is used for segmenting the tumor. Later, shape features and texture features have been extracted and classified using KNN and Ensemble Decision Tree model classifiers. The approach produced more false positive results. Authors in [12] have introduced radiomics-based machine learning classification algorithms. The experimental process has been conducted on the Breast Ultra-Sound Images (BUSI) Dataset. The authors in [13] have developed a computerized method that uses a combination thresholding approach to identify and extract the Breast Tumor Section (BTS) from 2D MRI slices. Based on Shannon's Entropy, Watershed Segmentation, and the Slime Mould Algorithm, a tri-level thresholding has been developed. The authors in [14] used the Extreme Learning Machine (ELM) to diagnose Breast Cancer. Irrelevant features have been removed using the gain ratio feature selection method. Using the Wisconsin Diagnostic Breast Cancer dataset, a cloud computing-based method for remote Breast Cancer diagnoses has been presented and tested. Breast thermal imaging is used to develop automated Breast Cancer diagnosis framework by authors in [15]. Initially, images

of various breast orientations have been used. Using image processing techniques, healthy image patches have been extracted and processed. Marine Predators technique has been used to extract features and the Decision Tree (DT) was utilized for classification. A layer connectivity-based architecture has been presented by authors in [16] for the segmentation of low contrast nodules from ultrasound images. The authors have blended high-level coarse segmentation with dense connectivity. The nodule has been refined using the dilated filter, and the architecture's accuracy has been increased by using the class imbalance loss function. The approach resulted in true negative results. The authors in [17] have presented a study that aims to analyse and predict Breast Cancer using a multi-model classification approach however, resulted in low accuracy. The authors in [18] have implemented four various types of feature selection methods, namely Bayesian Model Average (BMA) [19], Fast Backward Variable Selection (FastBw) [20], elastic net [21], and clinical expertise [22]. BMA approach use Best Subset Selection and regression analysis to choose the best model. However, BMA is applicable up to 30-40 features due to the overload on computation [23]. FastBw is applicable for more than 30-40 features. The FastBw method use multiple regression approach for the model containing all features and removes one feature at a time. This iteration process continues till it satisfies threshold or stopping rule [24]. Elastic net uses the linear combination of Lasso [25] and Ridge regression [26]. The approach finds the ridge regression coefficients initially and then proceeds with a lasso type of shrinkage [27]. The approach resulted in computation complexity issue. The authors in [28] proposed Two object detection models, YOLOv5 and Mask R-CNN. YOLOv5 is used to analyse the mass characteristics to distinguish between benign or malignant. However, YOLOV5's had limitations for certain tweaks to the original model to achieve the desired effects. Borders and size of tumors are detected using Mask RCNN to identify malignancies. The proposed model is trained on the INbreast, CBIS-DDSM. It is observed from the above comparative results that along with strong classifiers set of relevant features are very needed to achieve the better classification of breast tumors.

It is noticed that the most of the literatures above mentioned have limitation in classifying the breast tumors accurately due to True negative and False positive results. Many approaches are compatible with small size datasets. Some approaches ended up in computation complexity issues. Thus, the paper intends to overcome the challenge of early detection of Breast Cancer and it address the above issues by proposing the fusion scheme of the low-level features. The proposed framework use Mammogram images for analysing the morphological and texture features of the masses for classifying Breast tumors. The masses are segmented and localized using well-known scheme. Morphological features such as the shape and marginal characteristics of the mass are considered. However, these characteristics alone may not be sufficient to improve the classification accuracy [17]. Thus, the texture features of the mass were also extracted to analyse the pixel variation characteristics of the masses. The variance is considered as important parameters for fusing the features. The inter and intra group variances are calculated between features. While the variance is beyond certain threshold, the feature value is considered for the analysis or otherwise not. As a result, a fused feature is constructed with lower dimension for improving the quality of diagnosis. The proposed approach is explained in detail in the below section.
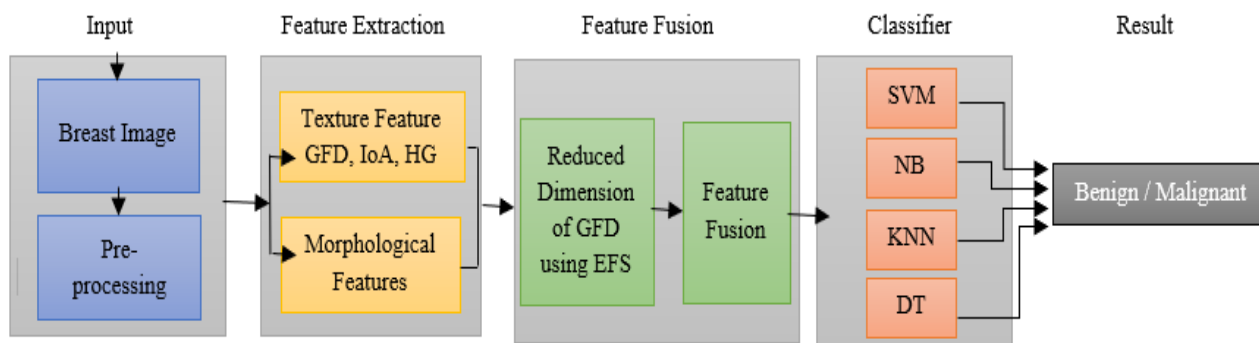

Figure. 1 Architecture view of the proposed approach

3. Proposed approach

3.1 Architecture view of morphological and texture features fusion

In this Section, the procedure for fusing the morphological and texture features is presented and explained. The proposed approach has used INbreast datasets (https://biokeanos.com/source/INBreast) [29] and CBIS-DDSM (https://complexity.cecs.ucf.edu/cbis-ddsm/) [30] for its experimentation. Fig. 1 depicts the flow diagram for fusing the feature and classification of masses either as benign or malignant. Input image is preprocessed for making it suitable for feature extraction. In this work, we extract both morphological and texture features. Initially, these features are linearly combined and considered as single feature. The combined feature is logically reduced based on the properties of inter and intra group variance.

3.2 Procedure to extract morphological features of the mass

The proposed approach considers the below morphological features of the mass and are mathematically represented in Eqs. (1)-(16). These morphological features are considered as a measure for tumor classification. The more information on the implementation and experiments are available in [17].

Morphological Features
The *Area* (*A*) of the mass is captured through the number of pixels and is represented in Eq. (1). The *p* and *q* are representing the rows and columns of the boundary of the RoI. Let K_{mass} and K_{margin} represents the RoI and boundary of the RoI in the image respectively. $$Area(A) = \sum_{i=1}^{p} \sum_{j=1}^{q} c[K_{mass}(i,j)] = 1 \quad (1)$$
The *Perimeter* (*P*) of the mass represents the number of pixels in the boundary of mass and is represented in Eq. (2). The *p* and *q* represent the row and column of the boundary of the mass. $$Perimeter(P) = \sum_{i=1}^{p} \sum_{j=1}^{q} c[K_{margin}(i,j)] = 1 \quad (2)$$
The *Distance* (*D*) represents these distances between the centroid and a pixel in the boundary and is depicted in Eq. (3). The (a_1, b_1) and (a_2, b_2) represents the co-ordinates of centroid and

boundary pixels respectively.

$$Distance\ (D) = \sqrt{(b_2 - b_1)^2 + (a_2 - a_1)^2} \quad (3)$$

The *Bending Energy* (B_E) defines the curvature features of mass/shape of the object and is mathematically represented in Eq. (4). Here, *C(s)* is the curvature function, *s* is length of arc and *N* represents number of pixels in boundary of the masses.

$$Binding\ Energy\ (B_E) = \frac{1}{N}\sum_{s=0}^{N-1} C(s)^2 \quad (4)$$

The *Eccentricity* (E_C) represents the aspect ratio and captures the elliptical property of the masses as shown in Eq. (5). The *major axis* captures the longest diameter of ellipse and *minor axis* captures shortest diameter, respectively.

$$Eccentricity\ (E_C) = \sqrt{1 - \left(\frac{minor\ axis^2}{major\ axis^2}\right)} \quad (5)$$

The *Circular Mass* (C_M) is a measure of roundness of the tumor to discriminate circular and elliptical property of the mass. This is mathematically represented in Eq. (6) where *CP* is Convex Perimeter.

$$Circular\ Mass\ (C_M) = \frac{4 \times \pi \times Area}{(CP)^2} \quad (6)$$

The circularity with respect to *Elliptical Mass* (E_M) is computed as depicted in Eq. (7). Maximum Radius (Max_R) captures maximum distance and Minimum Radius (Min_R) captures the minimum distance between centroid and boundary of mass.

$$Elliptical\ Mass\ (E_M) = \sqrt{\left(\frac{Min_R}{Max_R}\right)} \quad (7)$$

The *Elongatedness* (E_L) represents the ratio of mass area and the maximum mass thickness and is shown below in Eq. (8). Here, *t* is the maximum mass thickness without holes in mass.

$$Elongatedness\ (E_L) = \left[\frac{Area}{(2 \times t)^2}\right] \quad (8)$$

The *Equivalent Diameter* (E_D) defines the circle diameter and is represented in Eq. (9).

$$Equivalent\ Diameter(E_D) = \sqrt{4 \times \left(Area/\pi\right)} \quad (9)$$

686

The *Compactness* (C_P) defines the ratio of area of mass of a circle with its perimeters and is depicted in Eq. (10).

$$Compactness\ (C_P) = \left[\frac{(4 \times \pi \times Area)}{(Perimeter)^2}\right] \tag{10}$$

The *Solidity* (*S*) captures the density of mass and is measured as given below in Eq. (11).

$$Solidity\ (S) = \left[\frac{Area}{Covex\ Area}\right] \tag{11}$$

The *Convexity* (C_X) presents the tumor boundaries using convex perimeters and is represented in Eq. (12).

$$Convexity\ (C_X) = \left[\frac{Convex\ Perimeter}{Perimeter}\right] \tag{12}$$

The *Standard Deviation* refers to the scattering of data with respect to mean (μ) and is represented in Eq. (13).

$$(\sigma_{mass}/\sigma_{margin}) = \left[\sqrt{\frac{1}{N}\Sigma_1^N|x_i - \mu|^2}\right] \tag{13}$$

The *Shape Index* (S_I) defines the ratio of the perimeter of the mass and Maximum radius, and is represented in Eq. (14).

$$Shape\ Index\ (S_I) = \left[\frac{Perimeter}{2 \times Max_R}\right] \tag{14}$$

The *Entropy* (*E*) characterize the shape features in terms of the disorder, randomness of masses and is mathematically represented as given in below Eq. (15). Here, h_i represents normalized histogram counts.

$$Entropy\ (E) = \sum[h_i \times log_2(h_i)] \tag{15}$$

The *Skewness* (S_K) characterizes the data distribution of mass and is represented in Eq. (16). Here, *E(q)* represents the Expected value of quantity *Q*.

$$Skewness\ (S_K) = \left[\frac{E(Q-\mu)^3}{\sigma^3}\right] \tag{16}$$

Further, it is noticed that, texture features also play vital role in classifying the breast cells into benign and malignant. Hence, the proposed approach extracted the texture features to analyse and the process of extraction is given in the section 3.3.

3.3 Procedure to extract texture features of the mass

The texture is one of the important low-level features present in an image and proposed approach use it for classifying the RoI/ mass as benign or malignant. In this sub-section, the procedure for extracting the texture from RoI is presented along with theory of *Goodness-of-Fit Distribution (GFD), Independence of Attributes (IoA)* and *Homogeneity (HG)*. Initially, all these features are combined linearly and represented as single feature. The pre-processed RoI of mammogram image is given as input. The RoI of a mammogram is mathematically represented in Eq. (17).

$$I_{RoI}^{A \times B} = x(i,j) \tag{17}$$

Here, *x(i, j)* returns the intensity of a pixel at (*i, j*) *and A X B* is the size of the RoI. In general, the RoI may not always be in a perfect shape, such as square, rectangle, etc. However, during the segmentation process, we extract the RoI/mass and super impose it with a square such that processing the RoI is seamless. The procedure to bind the RoI/mass is depicted in Fig. 2. It is known that the pixels contribute to the RoI/mass and the feature is extracted from the characteristics of these contributing pixels. As a result, it is imperative that we process only those pixels that are part of the mass.

Mathematically, this can be represented as a predicate, which is a Boolean valued function as given in Eq. (18) and is called as a predicate on $I_{RoI}^{A \times B}$

$$P: I_{RoI}^{A \times B} \rightarrow \{TRUE, false\} \tag{18}$$

A predicate is given in Eq. (19) to determine whether *x (i, j)* falls within the RoI or outside RoI.

$$Predicate: I_{RoI}^{A \times B}(i,j) \rightarrow \{1,0\} \tag{19}$$

The texture feature captures the variations in the intensity/colour, contrast, chrominance and luminance between two neighbouring pixels in *vertical, diagonal,* and *horizontal* directions.
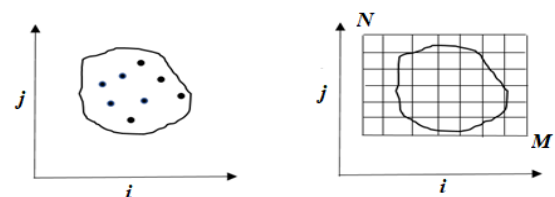
(a) Segmenting RoI   (b) Bounding with square
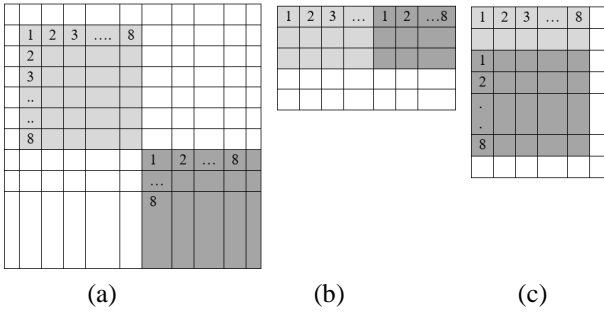
Figure. 2 RoI/Mass from mammogram

(a)   (b)   (c)

Figure. 3 *a X b* blocks of pixels in various directions: (a) Pass in *diagonal* dir, (b) Pass in *horizontal* dir, and (c) Pass in *vertical* dir
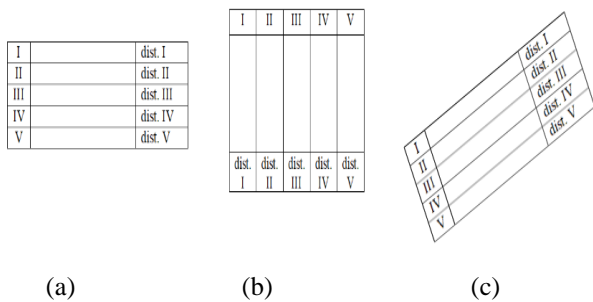

(a)   (b)   (c)

Figure. 4 Distribution of average pixel values in different directions: (a) *Horizontal* direction, (b) *Vertical* direction, and (c) *Diagonal* direction

The proposed approach has extracted the texture information using the theory of non-parametric testing on pixels in *horizontal, vertical* and *diagonal* directions. The predicate defined in Eq. (19) is extended in *horizontal, vertical* and *diagonal* directions and is given in Eqs. (20)-(22).

$$I_{RoI}^{A\times B}[x(i,j+1)] = P: I_{RoI}^{A\times B}(i,j+1) = 1 \quad (20)$$

$$I_{RoI}^{A\times B}[x(i+1,j)] = P: I_{RoI}^{A\times B}(i+1,j) = 1 \quad (21)$$

$$I_{RoI}^{A\times B}[x(i+1,j+1)] = P: I_{RoI}^{A\times B}(i+1,j+1) = 1 \quad (22)$$

The Eqs. (20)-(22) should satisfy Eq. (19) such that only those pixels that are part of the masses are considered. In Eqs. (20)-(22), $x(i, j+1)$, $x(i+1, j)$ and $x(i+1, j+1)$ represents the neighbouring pixels in *vertical, horizontal,* and *diagonal* directions. The proposed approach extract parameters such as *GFD, IoA* and *HG* and consider these as texture features. These values are extracted from $m X n$ blocks of RoI and will have an averaging effect with a value of $a=8$ and $b=8$. Thus, in the *vertical, horizontal* and *diagonal* directions, the trace is 8 x 8 pixels as shown in Fig. 3 below. The 8 x 8 window in *horizontal,*

vertical, and *diagonal* directions and the trace is one-pixel window is depicted. Thus, all the pixels in these directions contribute in measuring the texture. As a result, Eqs. (20)-(22) can be rewritten and represented in Eqs. (23)-(25).

$$I_{RoI}^{A\times B} = x^{a\times b}(i,j+1)|P[I_{RoI}^{a\times b}(i,j+1)] = 1 \quad (23)$$

$$I_{RoI}^{A\times B} = x^{a\times b}(i+1,j)|P[I_{RoI}^{a\times b}(i+1,j)] = 1 \quad (24)$$

$$I_{RoI}^{A\times B} = x^{a\times b}(i+1,j+1)|P[I_{RoI}^{a\times b}(i+1,j+1)] = 1 \quad (25)$$

The grey values in each cell of blocks are replaced with a threshold function as shown below in Eq. (26).

$$Grey_{Avg} = \sum_{i,j\in a\times b}^{a,b}[x(i,j)/(a\times b)] \quad (26)$$

$$x_{i\in a,j\in b}^{a\times b}(i,j) = \begin{cases} r_5 & if\ 128 \leq Grey_{Avg} \leq 255 \\ r_4 & 64 \leq Grey_{Avg} \leq 127 \\ r_3 & 32 \leq Grey_{Avg} \leq 63 \\ r_2 & 0 \leq Grey_{Avg} \leq 31 \\ r_1 & Otherwise \end{cases} \quad (27)$$

In the above Eq. (27), *r1=0, r2=32, r3=64, r4=128* and *r5=255* and the cell values of any block is a grey value. The purpose of averaging is to enable the non-parametric test, since it considers different distributions for testing. In the proposed approach, we consider the value of $Grey_{Avg}$ as a distribution such that the test can estimate the presence of texture information. The procedure for measuring these parameters is presented in below sub-section 3.3.1.

3.3.1 Measuring texture features

As discussed earlier, the texture component of RoI/mass is extracted by applying a non-parametric test on the averaged pixel values. The non-parametric test provides information on the relationship between two or more samples. The test is conducted on averaged pixel values to estimate the presence of texture features through *GFD, IoA,* and *HG*. The average pixel values are distributed in various directions such as *horizontal, vertical* and *diagonal* as shown in Fig. 4. The *GFD* is calculated using $GFD_{Avg}^{Direction}$ between the distributions. The *GFD* for each direction is represented in Eqs. (28)-(30). This is depicted in Fig. 4.

$$GFD_{Avg}^{Horizontal} = \left(\frac{\left(\sum_i^n GFD_i^{Hori}\right)}{n} \right) \quad (28)$$

$$GFD_{Avg}^{Vertical} = \left(\frac{\left(\sum_i^n GFD_i^{Vert}\right)}{n}\right) \tag{29}$$

$$GFD_{Avg}^{Diagnol} = \left(\frac{\left(\sum_i^n GFD_i^{Diag}\right)}{n}\right) \tag{30}$$

In Eqs. (28)-(30), $GFD_i^{Direction}$ is calculated for any i^{th} 8x8 block in *horizontal, vertical* and *diagonal* directions. Similarly, the variance of $GFD_{var}^{Direction}$ can be calculated using Eq. (31).

$$GFD_{Var}^{Direction} = \left(\sum_{i=1}^n \frac{\left(GFD_i^{Direc} - GFD_{Avg}^{Direc}\right)}{n-1}\right) \tag{31}$$

In Eq. (31), GFD_i^{Direc} represents the *GFD* of any i^{th} 8x8 block in a direction (*vertical, horizontal* and *diagonal*) and GFD_{Avg}^{Direc} is the average *GFD* calculated in a direction (*vertical, horizontal* and *diagonal*). The *IoA* for each direction is calculated using Eqs. (32)-(34).

$$IoA_{Avg}^{Horizontal} = \sum_{i=1}^n \left(\frac{\left(o_i^{Hori} - e_i^{Hori}\right)}{e_i^{Hori}}\right) \tag{32}$$

$$IoA_{Avg}^{Vertical} = \sum_{i=1}^n \left(\frac{\left(o_i^{Vert} - e_i^{Vert}\right)}{e_i^{Vert}}\right) \tag{33}$$

$$IoA_{Avg}^{diagnol} = \sum_i^n \left(\frac{\left(o_i^{diag} - e_i^{diag}\right)}{e_i^{diag}}\right) \tag{34}$$

In Eq. (32)-(34), o_i^{Hori}, o_i^{Vert} and o_i^{diag} is the observed frequency in *horizontal, vertical* and *diagonal* blocks. Similarly, e_i^{Hori}, e_i^{Vert} and e_i^{diag} are the expected frequency in the *horizontal, vertical* and *diagonal* directions respectively. The variance of $IoA_i^{direction}$ is calculated using Eq. (35) below.

$$IoA_{Var}^{Direction} = \sum_{i=1}^n \frac{\left(IoA_i^{Direc} - IoA_{Avg}^{Direc}\right)}{n-1} \tag{35}$$

In Eq. (36), IoA_i^{Direc} is the Independence of Attributes of any i^{th} 8x8 block in *horizontal, vertical* and *diagonal* directions. The IoA_{Avg}^{Direc} is the Average of Independence of Attributes for the respective 8x8 block in *horizontal, vertical* and *diagonal* directions. The *HG* is measured using the nature of the distribution of the pixels in RoI, such as Binomial, Poisson, and Normal distributions. This means, based on the trend followed by the distribution of average pixel values in *horizontal, vertical,* and *diagonal* directions, the homogenous nature of the average

pixel values can be measured. In terms of Binomial distribution, the *HG* is represented in Eq. (36) below.

$$P\left(HG_{Vert,hori,diag \in Directions}^{Direc}Thresh: n, p\right) =$$
$$nC_{HG_{Thresh}^{Direc}} HG_{Thresh}^{Direc} \times p^{HG_{Thresh}^{Direc}} \times (1-P)^{HG_{Thresh}^{Direc}} \tag{36}$$

In the above Eq. (36), HG_{Thresh}^{Direc} is the count of HG_i^{Direc} such that HG_i^{Direc} is greater than HG_{Thresh}^{Direc} more than $n/2$ times. The $n*p$ gives the mean Binomial distribution of HG_{Threh}^{Direc} and the variance of Binomial distribution of HG_i^{Direc} is $n*p(1-p)$. The *HG* is characterized based on skewness. While $p=0.5$, the distribution HG_i^{Direc} is symmetric around the mean and the presence of texture is low. While $p<0.5$ and $p>0.5$, the distribution HG_i^{Direc} is left and right skewed. As a result, the presence of the texture is estimated based on the value of *p*. Similarly, the texture content is also measured based on the Poisson distribution characteristics of the samples. This is represented in Eq. (37) below.

$$F\left(HG_{Thres}^{Direc}: \lambda\right) = \Pr\left(HG_i^{Direc} = HG_{Thres}^{Direc}\right) =$$
$$\left(\frac{\lambda^{HG_i^{Direc}} \times e^{-\lambda}}{HG_{Thresh}^{Direc}}\right) \tag{37}$$

In the above Eq. (37), *e* is the Euler's number (e=2.71828). The positive real number $\boldsymbol{\lambda}$ is equal to the expected value of HG_i^{Direc} and also is variance. The variance measured through below Eq. (38) gives the texture content.

$$\lambda = Eu\left(HG_i^{Direc}\right) = Var(HG_i^{Direc}) \tag{38}$$

Similarly, if the distribution of HG_{Thres}^{Direc} is normal, texture characteristics is estimated accordingly. After analyzing and estimating the importance of morphological and textural features, it is predicted that the fusion of both of these futures plays a vital role in achieving the accurate classification of Breast Cancer cells. The process of feature fusion is explained detail in below section.

3.4 Procedure to fuse both morphological and texture features

In this subsection, the morphological features presented in Section 3.1 and texture features presented in Section 3.2 are fused. It is found that the characteristics and discrimination capability of each type of features on mammogram mass are encouraging. For example, the morphological

features can be effective in discriminating a set of masses and similarly, the texture feature may be effective for the different categories of masses. Thus, these features can be fused as single features for improving the accuracy in classifying malignant from benign masses. The procedure for fusing the feature is explained below.

Let F_{morp} be the morphological feature of the mass and is denoted in Eq. (39), where n denotes the size of the morphological feature,

$$F_{morp} = \{f_{morp1}, f_{morp2}, f_{morpi}, \ldots, f_{morpn}\} \quad (39)$$

Let F_{GFD} is *Goodness of Fit distribution* (*GFD*) of the mass in all the directions and is denoted in Eq. (40), where m is the dimension of F_{GFD}.

$$F_{GFD} = \{f_{GFD1}, f_{GFD2}, f_{GFDi}, \ldots, f_{GFDm}\} \quad (40)$$

Let F_{IoA} is *Independence of Attributes* (*IoA*) of the mass in all the directions and is shown in Eq. (41), where p is the dimension of F_{IoA}.

$$F_{IoA} = \{f_{31}, f_{32}, f_{3i}, \ldots, f_{3n}\} = \{f_{IoA1}, f_{IoA2}, f_{IoAi}, \ldots, f_{IoAp}\} \quad (41)$$

Let F_{HG} is *Homogeneity* (*HG*) of the mass in all the directions and is represented in Eq. (42), where q is the dimension of F_{HG}.

$$F_{HG} = \{f_{HG1}, f_{HG2}, f_{HGi}, \ldots, f_{HGq}\} \quad (42)$$

As a initial step, these four features are considered as a single feature by linearly combining all of them, and mathematically it is represented in Eq. (43)

$$F = \{F_{morp}, F_{GFD}, F_{IoA}, F_{HG}\} \quad (43)$$

However, combining the features linearly increases dimension and thus there is a curse of dimensionality. Thus, it is imperative that a feature subset from each feature is required to further improve the classification accuracy.

In this approach, variance is used as a measure for fusing the features effectively. The process requires the size of the feature should be same in each factor level. As a result, the dimension of F_{morp} is reduced to match the dimension of F_{GFD}, F_{IoA} and F_{HG} using Exhaustive Feature Selection (EFS) techniques. The dimension of F_{morp} is n and EFS strategy produces 2^n-1 feature combinations. The proposed approach considers all the feature combinations having its size equal to m and among them which ever provides
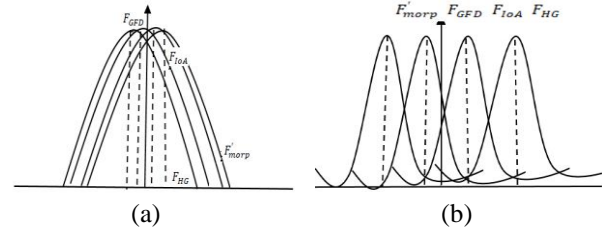

Figure 5. The F value on features $(F'_{morp}, F_{GFD}, F_{IoA}$ and $F_{HG})$: (a) Low F value and (b) High F value

higher performance is selected and is denoted as F'_{morp}.

Each mass in a mammogram is considered as RoI and all the features say, $F_{morp}, F_{GFD}, F_{IoA}$ and F_{HG} are extracted. The F_{morp} is decomposed to F'_{morp} using EFS for making the dimension of all the features equal (i.e, $n=m=p=q$). The algorithm exploits the variance among the features and within the features (i.e variance between the group and variance among the group).

It is observed that the distribution of features in Fig. 5(a) is similar and thus the F-value is low. Thus, all the features say $F'_{morp}, F_{GFD}, F_{IoA}$ and F_{HG} are having similar properties in discriminating the masses. In contrast, the F-value of distribution shown in Fig. 5(b) is high, each distribution is different and thus more discriminating power. Thus, the F-value is calculated using Eqs. (39)-(43). If the F-value is less than a threshold, T, elements in Feature are removed till the F value is good. Within Group Variance (*WGV*) and Between Group Variance (*BGV*) are playing critical role in the algorithm and both of them are calculated as given below.

Within Group Variance (*WGV*) for each feature is calculated as given in Eq. (44), where, $y_{ij}^{F'_{morp}}$ and $\bar{y}_j^{F'_{morp}}$ are the dimension reduced morphological features (F'_{morp}) and mean of the F'_{morp} respectively.

$$WGV(F'_{morp}) = \sum_{j=1}^{F'_{morp}} \sum_{i=1}^{n} \left(y_{ij}^{F'_{morp}} - \bar{y}_j^{F'_{morp}}\right)^2 \quad (44)$$

Similarly, Within Group Variance (*WGV*) for F_{GFD}, F_{IoA} and F_{HG} are calculated as shown in Eq. (45)-(47).

$$WGV(F_{GFD}) = \sum_{j=1}^{F_{GFD}} \sum_{i=1}^{n} \left(y_{ij}^{F_{GFD}} - \bar{y}_j^{F_{GFD}}\right)^2 \quad (45)$$

In the above Eq. (45), $y_{ij}^{F_{GFD}}$ and $\bar{y}_j^{F_{GFD}}$ are the values in F_{GFD} and their mean respectively and n is the dimension of F_{GFD}.

$$WGV(F_{IoA}) = \sum_{j=1}^{F_{IoA}} \sum_{i=1}^{n} \left(y_{ij}^{F_{IoA}} - \bar{y}_j^{F_{IoA}} \right)^2 \quad (46)$$

In the above Eq. (46), $y_{ij}^{F_{IoA}}$ and $\bar{y}_j^{F_{IoA}}$ are the values in F_{IoA} and their mean respectively and n is the dimension of F_{IoA}.

$$WGV(F_{HG}) = \sum_{j=1}^{F_{HG}} \sum_{i=1}^{n} \left(y_{ij}^{F_{HG}} - \bar{y}_j^{F_{HG}} \right)^2 \quad (47)$$

In the above Eq. (47), $y_{ij}^{F_{HG}}$ and $\bar{y}_j^{F_{HG}}$ are the values in F_{HG} and their mean respectively and n is the dimension of F_{HG}.

Similarly, Between Group Variance (BGV) is also calculated as shown in Eq. (48).

$$BGV(F'_{morp}, F_{GFD}, F_{IoA}, F_{HG}) = \sum nx(\bar{x} - \bar{x}_{GM})^2 \quad (48)$$

In the above Eq. (48), $\bar{x}$ is the sample mean (mean of $F'_{morp}, F_{GFD}, F_{IoA}$ and F_{HG}) respectively, $\bar{x}_{GM}$ is the group mean of all the features and n is the number of samples in all the features. The F-value is the ratio of BGV and WGV and is derived from Eqs. (45)-(47) and Eq. (48).

4. Experimental results

The performance of the proposed approach is evaluated on two mammographic datasets namely INbreast and CBIS-DDSM. The INbreast dataset has 7632 mammographic images and they are considered after data augmentation followed by adaptive histogram normalizations. Out of which, 2520 images are benign and 5112 images are malignant tumors. The Digital Database for Screening Mammography (DDSM) is a mammographic image dataset and it consists of 2620 scanned films with verified pathology information. It includes normal, malignant, and benign tissues. The CBIS-DDSM is a standardized and updated version of the DDSM dataset. The details, such as the number of malignant, benign, and benign without call-back masses are presented in Table 1. The CBIS-DDSM dataset is larger and INbreast has fewer samples. The training and testing splits of INbreast, CBIS-DDSM calc and CBIS-DDSM mass datasets are shown in Table 2. The proposed approach has used 6106, 1014, and 1228 mammogram images as training set from INbreast, CBIS-DDSM Calc, and CBIS-DDSM mass datasets to train the SVM classifier. We have used

Table 1. Details of CBIS-DDSM and INbreast Mammogram Datasets

Dataset	Data type	Malignant	Benign	Benign without call-back
CBIS-DDSM	Mass-Training RoI and Cropped Images	650	578	104
	Mass-Testing RoI and Cropped Images	147	195	37
	Calc-Training RoI and Cropped Images	544	528	474
	Calc-Testing RoI and Cropped Images	129	130	67
INbreast	Training	2045	1008	-
	Testing	511	252	-

Table 2: Training and Testing sets from INbreast, CBIS-DDSM calc and CBIS-DDSM mass datasets

Dataset	Split classes	Benign	Malignant	Total
INbreast	Training set	2016	4090	6106
	Testing set	504	1022	1526
CBIS-DDSM	Calc-Training RoI and Cropped Images	441	573	1014
	Calc-Testing RoI and Cropped Images	125	129	254
	Mass-Training RoI and Cropped Images	578	650	1228
	Mass-Testing RoI and Cropped Images	195	147	342

Table 3 Classification accuracy on INbreast using fused features

Classified as Benign		Classified as Malignant		Total	Classification Accuracy (%)	
SVM	ANN	SVM	ANN		SVM	ANN
454	492	50	12	504	90.06	97.61
82	51	940	971	1022	91.97	95.00
Overall				**1526**	**91.01**	**98.56**

Table 4 Classification accuracy CBIS-DDSM using fused features

Mass	Classified as Benign		Classified as Malignant		Total	Accuracy (%)	
	SVM	ANN	SVM	ANN		SVM	ANN
Calc RoI and Cropped Images	116	119	9	6	125	92.80	95.20
	4	3	125	126	129	96.89	97.67
	Overall				**254**	**94.84**	**96.43**
Mass- RoI and Cropped Images	183	181	12	14	195	93.84	92.82
	6	4	141	143	147	95.91	97.27
	Overall				**342**	**94.87**	**95.04**

Table 5 Precision, Recall, F-measure, and Accuracy on fused feature

Evaluation Metrics	Performance Measure					
	INbreast Dataset		CBIS-DDSM Dataset			
	SVM (%)	ANN (%)	SVM (%)		ANN (%)	
			Calc RoI	Mass- RoI	Calc RoI	Mass- RoI
Recall	89.40	88.21	88.36	90.18	90.33	86.26
Precision	92.72	95.76	92.31	93.10	94.31	90.03
F-Score	91.08	91.81	90.29	91.57	92.27	90.95
Accuracy	**91.01**	**98.56**	**94.84**	**94.87**	**96.43**	**95.04**

1526, 254 and 342 mammogram images as test images from INbreast, CBIS-DDSM Calc, and CBIS-DDSM mass datasets to evaluate the performance of proposed texture features.

This section discusses the performance evaluation of fused feature. The experiments are performed using MATLAB2020b on a Desktop Computer Core i7 having 8GB of graphics card and 16GB RAM. The fused features along with SVM, KNN and DT are used for classification. The proposed approach chosen training/testing ratio 70:30, 60:40, 50:50 to validate the fused feature. The cross-validation value is selected at 10 for all experiments. The performance is evaluated using the following statistical measures such as True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN). This is depicted in Table 3 and 4. On the INbreast dataset, the SVM classifies 454 mammograms as benign and ANN classifies 492 mammograms as benign on 504 test case mammogram images. SVM classifies 940 mammograms as malignant and ANN classifier classifies 971 mammograms as benign from 1022 test case mammogram images. The overall accuracy of the proposed feature on the INbreast dataset is 91%

for SVM and 98.56% for ANN. For the CBIS-DDSM Calc RoI dataset, the SVM classifier classifies 116 mammograms as benign and ANN classifies 119 mammograms as benign on 125 mammogram images. SVM classifies 125 mammograms as malignant and ANN classifies 126 mammograms as malignant from 129 test cases. The overall accuracy of the proposed feature for the CBIS-DDSM Calc RoI dataset is 94% for SVM and 96% for ANN. For the CBIS-DDSM Mass RoI dataset, the SVM classifies 183 mammograms as benign and ANN classifies 181 mammograms as benign from 195 mammogram images. SVM classifies 141 mammograms and ANN classifies 143 mammograms as malignant from 147 test cases. The overall accuracy of the proposed feature for the CBIS-DDSM Mass RoI dataset is 94% using SVM and 95% using ANN classifier.

In addition, Precision, Recall, F1-Score, and Accuracy are evaluated on fused feature and the results are presented in Table. 5. Table 6 presents the performance evaluation of the proposed fused feature with other comparative approaches on INbreast and CBIS-DDSM datasets.   The authors in [8] have adopted Hybrid thresholding and the machine

Table 6. Performance comparison of the fused feature.

Comp. Approaches	Features	Datasets	Classifiers	Acc (%)
Proposed Approach	Fused Feature	INbreast	SVM	91.01
			ANN	**98.56**
		CBIS-DDSM	SVM	94.85
			ANN	**95.73**
Anas and Haq (2024) [28]	Multi features	CBIS-DDSM and INbreast	YOLOv5+Mask RCNN	92.02
Teng and Zhang (2022) [18]	<30-40 Multi-features	INbreast,	Bayesian Model Average (BMA)	80.2
Teng and Zhang (2022) [18]	>30-40 Multi-features	INbreast	Fast Backward Variable Selection (FastBw)	79.9
Teng and Zhang (2022) [18]	>30-40 Multi-features	INbreast	elastic net	79.9
Zebari and Ibrahim (2021) [8]	Multi Features-50 features	DDSM INbreast,	ANN	81.52 82.21
Vaka and Soni, (2020) [31]	Fusion approach	DDSM	Histo-sigmoid fuzzy clustering	93.52
Zhang and Wu, (2020) [10]	Fusion approach	INbreast DDSM	Adaboost	86.20 91.12
Jiang and Xu, (2017) [32]	Histogram statistical features and GLCM	DDSM	Random Forest (RF)	77.05
Uthoff, and Sieren, (2018) [33]	13 histogram, texture, 18 shape	DDSM	ANN	95.23
Pashoutan and Shokouhi, (2017) [34]	Intensity, shape, texture	DDSM	FGMM	93.51

learning techniques to locate the RoI to classify benign or malignant Breast Cancer from mammogram images. The noise of each block has been reduced by applying the wavelet transform using BayesShrink soft thresholding, which captures high and low frequencies within various sub-bands. The methodology has been assessed on MIAS, DDSM, INbreast, and BCDR mammography imaging datasets. The authors in [31] employed Entropy, texture, and geometric features for detecting Breast Cancer tumors. The authors adopted Gaussian filter to eliminate the noise and machine learning techniques are used for classification. A multi-feature fusion based model called DE-Ada* has been presented by authors in [10] for the classification of breast masses. Various techniques have been employed to categorize the mammograms such as Scale-Invariant Feature Transform (SIFT), Histogram Of Oriented Gradient (HOG), Local Binary Pattern (LBP), Residual Network (ResNet), Densely linked convolutional Networks (DenseNet), and Visual Geometry Group (VGG). Mohanty and Rup (2019) proposed a hybrid Computer Aided Diagnosis (CAD) framework that identifies suspicious regions as RoI and classified them as either abnormal or normal. Various classifiers such as

SVM, K-NN, Naïve Bayes, and C4.5 are adopted to classify the RoI as benign or malignant. The authors in [32] have extracted histogram statistical features and texture features using GLCM to classify Breast Cancer tumors. The authors in [33] extracted intensity, texture, and shape features from mammograms. The authors employed K-medoids clustering to remove intra-correlated features and thereby reducing feature set. The authors used information theory approach for feature selection. The authors in [34] have used wavelet transform, Gabor wavelet transform, Zernike moments and GLCM to segment the region of the tumor in mammography. P values are used to choose the relevant features and Multi-Layer Perceptron (MLP) is used to classify the malignant or benign breast masses. It is observed from Table 6 that the performance of the fused-feature is encouraging on both the datasets. The SVM classifier has achieved in 91.01% and 94.55% of accuracy on INbreast and CBIS-DDSM datasets. Similarly, the classification accuracy using ANN is 98.56% and 95.731% on INbreast and CBIS-DDSM datasets respectively. The authors in [33] have shown close accuracy 95.3% on DDSM dataset using ANN classifier. The approach in [31] and [34] have achieved the classification

accuracy in the range of 93% on DDSM datasets. The authors in [34] have classified the DDSM dataset masses with 77.05% classification accuracy. The authors in [28] achieved 92.02% accuracy. Overall, the proposed fused feature has performed well. This is due to the fact that it combines both morphological and texture effectively and both of these features are good in discriminating malignant and benign masses. The properties of inter and intra group variance of both the features have reduced the dimension of the resultant feature and has retained the characteristics of the tumor.

5. Conclusion

In this paper we have fused both morphological and texture feature. The morphological feature has extracted shape and marginal properties of the masses. The texture feature has captured the variations in pixels of masses through Goodness of Fit Distribution (GFD), Independence of Attributes (IoA) and Homogeneity (HG). These features are initially combined together and the dimension of the feature is reduced logically. The inter and intra group variances between the features are estimated and both the feature is fused accordingly. The evaluation of the fused feature is performed on DDSM and INbreast Dataset, and Precision, Recall, F1-Score and Accuracy are measured. The SVM and ANN are used as classifiers. The overall accuracy of the proposed feature on the INbreast dataset is 91% for SVM and 98.56% for ANN. The overall accuracy of the proposed feature for the CBIS-DDSM Calc RoI dataset is 94% for SVM and 96% for ANN. The overall accuracy of the proposed feature for the CBIS-DDSM Mass RoI dataset is 94% using SVM and 95% using ANN classifier. In addition, Precision, Recall, F1-Score are evaluated. The result is compared with contemporary methods and found that the fused feature is performing well. In future we will extract volume features of the mass and fused in with the proper one.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization and Methodology- Shaila SG and Vadivel A; Data curation Formal analysis and Visualization- Gurudas VR; Software and Resources - Gurudas VR; Supervision and Investigation- Shaila SG; Writing—original draft preparation-Gurudas VR; Validation, Review and Editing- Shaila SG and Vadivel A.

References

[1] https://www.who.int/news-room/fact-sheets/detail/breast-cancer

[2] A.K Suleiman, "Awareness and attitudes regarding Breast Cancer and breast self-examination among female Jordanian students", *Journal of basic and clinical pharmacy,* Vol. 5, No. 3. pp.74-8, 2014

[3] A.F. Moustafa, T.W. Cary, L.R. Sultan, S.M. Schultz, E.F. Conant, S.S. Venkatesh and C.M. Sehgal, "Color Doppler Ultrasound improves Machine Learning Diagnosis of Breast Cancer", *Diagnostics*, Vol. 10, 631, 2020.

[4] R.V. Abdel-Menon, P. Raha, S. Kothari, S. Chakraborty, I. Chakrabarti and R. Karim, "Automated Detection and Classification of Mass from Breast Ultrasound Images", In: *Proc. of Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, Patna, India, pp. 1-4, 2015.

[5] F.A. González-Luna, J. Hernández-López, and W. GomezFlores, "A Performance Evaluation of Machine Learning Techniques for Breast Ultrasound Classification", In: *Proc. of 16th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)*, Mexico City, Mexico, pp. 1-5, 2019.

[6] B. Mughal, M. Sharif, N. Muhammad and T. Saba, "A Novel Classification Scheme to decline the Mortality Rate among Women due to Breast Tumor", *Microscopy research and technique*, Vol. 81, No. 2, pp. 171-180, 2018.

[7] E. Michael, H. Ma, H. Li, F. Kulwa and J. Li, "Breast Cancer Segmentation Methods: Current Status and Future Potentials", *BioMed Research International*, pp. 1-29, 2021.

[8] D.A. Zebari, D.A. Ibrahim, D. Q. Zeebaree, H. Haron, M.S. Salih, R. Damaševičius and M.A. Mohammed, "Systematic Review of Computing Approaches for Breast Cancer Detection based Computer Aided Diagnosis using Mammogram Images", *Applied Artificial Intelligence*, Vol. 35, No. 15, pp. 2157-2203, 2021.

[9] V. Inamdar, S.G. Shaila and M.K. Singh, "FNAB-Based Prediction of Breast Cancer Category Using Evolutionary Programming Neural Ensemble", In: *Proc. of International Conference on Computational Vision and Bio-Inspired Computing,* Springer Singapore, Vol. 1318, pp. 653-663, 2021.

[10] H. Zhang, R. Wu, T. Yuan, Z. Jiang, S. Huang, J. Wu, J. Hua, Z. Niu and D. Ji, "DE-Ada*: A

Novel Model for Breast Mass Classification using Cross modal Pathological Semantic Mining and Organic Integration of Multi-feature Fusions", *Information Science*, Vol. 539, pp. 461-486, 2020.

[11] T. Sadad, A. Hussain, A. Munir, M. Habib, S. Ali Khan, S. Hussain, S. Yang and M. Alawairdhi, "Identification of Breast Malignancy by Marker-Controlled Watershed transformation and Hybrid Feature Set for Healthcare", *Applied Science*, Vol. 10, No. 6, pp. 1900, 2020.

[12] A.K. Mishra, P. Roy, S. Bandyopadhyay, and S.K. Das, "Breast Ultrasound Tumor Classification: A Machine Learning—Radiomics based Approach", *Expert System,* Vol. 38, 2021.

[13] S. Kadry, R. Damaševičius, D. Taniar, V. Rajinikanth and I.A. Lawal, "Extraction of Tumor in Breast MRI using Joint Thresholding and Segmentation - A Study", In: *Proc. of the Seventh International conference on Bio Signals, Images, and Instrumentation* (ICBSII), Chennai, India, pp. 1-5, 2021.

[14] V. Lahoura, H. Singh, A. Aggarwal, B Sharma, M Mohammed, R. Damaševičius, S Kadry, K Cengiz, "Cloud Computing-Based Framework for Breast Cancer Diagnosis Using Extreme Learning Machine", *Diagnostics,* Vol. 11, No. 2, pp. 241, 2021.

[15] V. Rajinikanth, S. Kadry, D. Taniar, R. Damaševičius and H. T. Rauf, "Breast-Cancer Detection using Thermal Images with Marine-Predators-Algorithm Selected Features", In: *Proc. of the Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII)*, Chennai, India, pp. 1-6, 2021.

[16] A. Ouahabi and A. Taleb-Ahmed, "Deep learning for real-time semantic segmentation: Application in ultrasound imaging", *Pattern Recognition Letters*, Vol. 144, No. 12, pp. 27-34, 2021.

[17] V.R. Gurudas, S.G. Shaila, A. Vadivel, "Breast Cancer Detection and Classification from Mammogram Images Using Multi-model Shape Features", *SN Computer Science*, Vol. 3, No. 5, pp. 404, 2022.

[18] J. Teng, H. Zhang, W. Liu, X-O Shu and F. YeA "Dynamic Bayesian Model for Breast Cancer Survival Prediction", *IEEE Journal of Biomedical and Health Informatics,* Vol. 26, No. 11, 2022.

[19] K. Y. Yeung, R. Bumgarner and A. Raftery, "Bayesian model averaging: Development of an improved multi-class, gene selection and classification tool for microarray data", *Bioinformatics*, Vol. 21, pp. 2394-2402, 2005.

[20] J. F. Lawless and K. Singhal, "Efficient screening of nonnormal regression models", *Biometrics*, Vol. 34, No. 2, pp. 318-327, 1978.

[21] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net", *Journal of the Royal Statistical Society Series B: Statistical Methodology*, Vol. 67, No. 2, pp. 301-320, 2005.

[22] R. Fan, Y. Chen, S. Nechuta, H. Cai, K Gu, L. Shi, P. Bao, Y. Shyr, X-O Shu, and F. Ye, "Prediction models for Breast Cancer prognosis among Asian women", *Cancer*, Vol. 127, No. 11, pp. 1758-1769, 2021.

[23] D. Fletcher, *Bayesian Model Averaging*, Berlin, Heidelberg, Germany: Springer, pp. 31-55, 2018.

[24] C. Porzelius, M. Schumacher, and H. Binder, "Sparse regression techniques in low-dimensional survival data settings", *Statistics and Computing*, Vol. 20, pp. 151-163, 2010.

[25] L. Torgo and J. Gama, "Regression by classification", In: *Proc. of the Advanced Artificial Intelligence*, pp. 51-60, 1996.

[26] T. Hastie, J. Taylor, R. Tibshirani and G. Walther, "Forward Stagewise Regression and the Monotone Lasso", *Electronic Journal of Statistics*, Vol. 1, pp. 1-29, 2007.

[27] M. H. R. Khan, A. Bhadra and T. Howlader, "Stability selection for lasso, ridge and elastic net implemented with AFT models", *Statistical Applications in Genetics and Molecular Biology*, Vol. 18, No. 5, 2019.

[28] M. Anas, I.U. Haq, G. Husnain and S.A.F. Jaffery, "Advancing Breast Cancer Detection: Enhancing YOLOv5 Network for Accurate Classification in Mammogram Images", *IEEE Access*, Vol. 12, pp. 16474-16488, 2024.

[29] https://biokeanos.com/source/INBreast

[30] https://complexity.cecs.ucf.edu/cbis-ddsm/

[31] A.R. Vaka, B. Soni and S.R. Reddy, "Breast Cancer detection by leveraging Machine Learning", *ICT Express*, Vol. 6, No. 4, pp. 320-324, 2020.

[32] Z. Jiang, W. Xu and S. Chen, "Classification of benign and malignant Breast Cancer based on DWI texture features", In: *Proc of the International Conference on Bioinformatics and Computational Intelligence*, pp. 11-15, 2017.

[33] J. Uthoff and J. C. Sieren, "Information theory optimization based feature selection in breast mammography lesion classification", In: *Proc. of IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, Washington, DC, USA, pp. 817-821, 2018.

[34] S. Pashoutan, S. Baradaran Shokouhi and M. Pashoutan, "Automatic Breast Tumor Classification Using a Level Set Method and Feature Extraction in Mammography", In: *Proc. of 24th National and 2nd International Iranian Conference on Biomedical Engineering (ICBME)*, Tehran, Iran, pp. 1-6, 2017.