Review Article
Involvement of Machine Learning for Breast Cancer Image Classification: A Survey

Abdullah-Al Nahid and Yinan Kong

School of Engineering, Macquarie University, Sydney, NSW 2109, Australia

Correspondence should be addressed to Abdullah-Al Nahid; abdullah-al.nahid@students.mq.edu.au

Breast cancer is one of the largest causes of women's death in the world today. Advance engineering of natural image classification techniques and Artificial Intelligence methods has largely been used for the breast-image classification task. The involvement of digital image classification allows the doctor and the physicians a second opinion, and it saves the doctors' and physicians' time. Despite the various publications on breast image classification, very few review papers are available which provide a detailed description of breast cancer image classification techniques, feature extraction and selection procedures, classification measuring parameterizations, and image classification findings. We have put a special emphasis on the Convolutional Neural Network (CNN) method for breast image classification. Along with the CNN method we have also described the involvement of the conventional Neural Network (NN), Logic Based classifiers such as the Random Forest (RF) algorithm, Support Vector Machines (SVM), Bayesian methods, and a few of the semisupervised and unsupervised methods which have been used for breast image classification.

1. Introduction

The cell of the body maintains a cycle of regeneration processes. The balanced growth and death rate of the cells normally maintain the natural working mechanism of the body, but this is not always the case. Sometimes an abnormal situation occurs, where a few cells may start growing aberrantly. This abnormal growth of cells creates cancer, which can start from any part of the body and be distributed to any other part. Different types of cancer can be formed in human body; among them breast cancer creates a serious health concern. Due to the anatomy of the human body, women are more vulnerable to breast cancer than men. Among the different reasons for breast cancer, age, family history, breast density, obesity, and alcohol intake are reasons for breast cancer.

Statistics reveal that in the recent past the situation has become worse. As a case study, Figure 1 shows the breast cancer situation in Australia for the last 12 years. This figure also shows the number of new males and females to start suffering from breast cancer. In 2007, the number of new cases

for breast cancer was 12775, while the expected number of new cancer patients in 2018 will be 18235. Statistics show that, in the last decade, the number of new cancer disease patients increased every year at an alarming rate.

Figure 2 shows the number of males and females facing death due to breast cancer. It is predicted that in 2018 around 3156 people will face death; among them 3128 will be women which is almost 99.11% of the overall deaths due to breast cancer.

Women's breasts are constructed by lobules, ducts, nipples, and fatty tissues. Milk is created in lobules and carried towards nipple by ducts. Normally epithelial tumors grow inside lobules as well as ducts and later form cancer inside the breast [1]. Once the cancer has started it also spreads to other parts of the body. Figure 3 shows the internal construction from a breast image.

Breast cancer tumors can be categorized into two broad scenarios.

(i) Benign (Noncancerous). Benign cases are considered as noncancerous, that is, non-life-threatening. But on a few
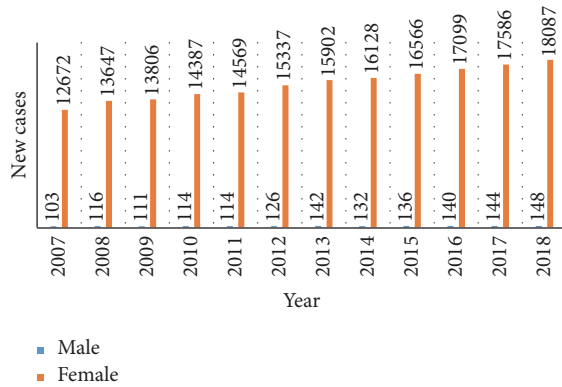
FIGURE 1: Number of new people facing cancer in Australia from 2007 to 2018 [5].
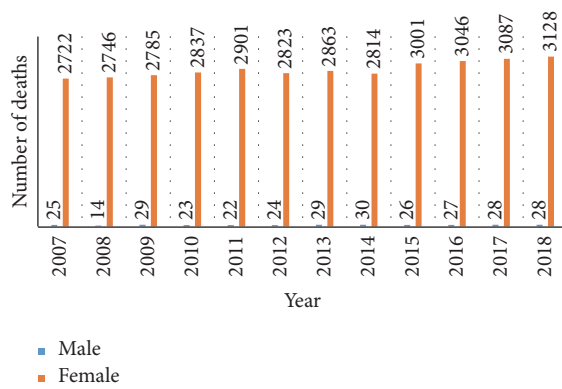

FIGURE 2: Number of people dying due to cancer in Australia from 2007 to 2018 [5].

occasions it could turn into a cancer status. An immune system known as "sac" normally segregates benign tumors from other cells and can be easily removed from the body.

(ii) Malignant (Cancerous). Malignant cancer starts from an abnormal cell growth and might rapidly spread or invade nearby tissue. Normally the nuclei of the malignant tissue are much bigger than in normal tissue, which can be life-threatening in future stages.

Cancer is always a life-threatening disease. Proper treatment of cancer saves people's lives. Identification of the normal, benign, and malignant tissues is a very important step for further treatment of cancer. For the identification of benign and malignant conditions, imaging of the targeted area of the body helps the doctor and the physician in further diagnosis. With the advanced modern photography techniques, the image of the targeted part of the body can be captured more reliably. Based on the penetration of the skin and damage of the tissue medical photography techniques can be classified into two groups.

(i) Noninvasive. (a) Ultrasound: this photography technique uses similar techniques to SOund Navigation And Ranging (SONAR) which operates in the very-high-frequency domain and records the echos of that frequency, invented by Karl

Theodore Dussik [2]. An ultrasound image machine contains a Central Processing Unit (CPU), transducer, a display unit, and a few other peripheral devices. This device is capable of capturing both 2D and 3D images. Ultrasound techniques do not have any side-effects, with some exceptions like production of heat bubbles around the targeted tissue. (b) X-ray: X-rays utilize electromagnetic radiation, invented by Wilhelm Conrad Roentgen in 1895. The mammogram is a special kind of X-ray (low-dose) imaging technique which is used to capture a detailed image of the breast [3]. X-rays sometimes increase the hydrogen peroxide level of the blood, which may cause cell damage. Sometimes X-rays may change the base of DNA. (c) Computer Aided Tomography (CAT): CAT, or in short CT imaging, is advanced engineering of X-ray imaging techniques, where the X-ray images are taken at different angles. The CT imaging technique was invented in 1970 and has been mostly used for three-dimensional imaging. (d) Magnetic Resonance Imaging (MRI): MRI is a noninvasive imaging technique which produces a 3D image of the body, invented by Professor Sir Peter Marsfield, and this method utilizes both a magnetic field as well as radio waves to capture the images [4]. MRI techniques take longer to capture images, which may create discomfort for the user. Extra cautions need to be addressed to patients who may have implanted extra metal.

(ii) Invasive. (a) Histopathological images (biopsy imaging): histopathology is the microscopic investigation of a tissue. For histopathological investigation, a patient needs to go through a number of surgical steps. The photographs taken from the histopathological tissue provide histopathological images (see Figure 4).

2. Breast Image Classification

Various algorithms and investigation methods have been used by researchers to investigate breast images from different perspectives depending on the demand of the disease, the status of the disease, and the quality of the images. Among the different tasks, for breast image classification, machine learning (ML) and the Artificial Intelligence (AI) are heavily utilized. A general breast image classifier consists of four stages (see Figure 5):

 (i) Selection of a breast database

 (ii) Feature extraction and selection

 (iii) Classifier model

 (iv) Performance measuring parameter

 (v) Classifier output.

Figure 5 shows a very basic breast image classifier model.

2.1. Available Breast Image Databases. Doctors and physicians are heavily reliant on the ultrasound, MRI, X-ray, and so forth images to find the breast cancer present status. However, to ease the doctors' work, some research groups are investigating how to use computers more reliably for breast cancer diagnostics. To make a reliable decision about the cancer
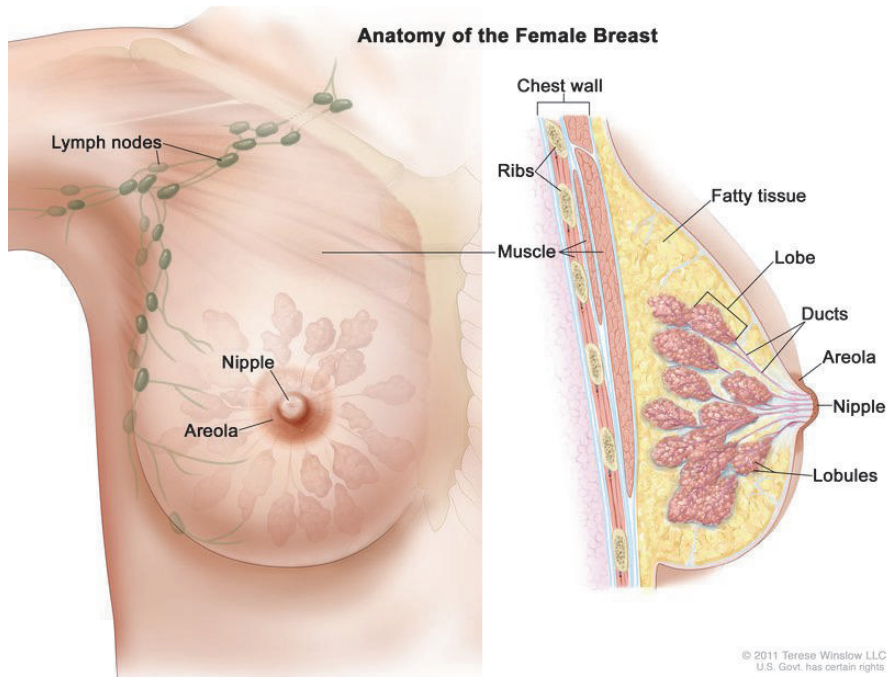
Figure 3: Anatomy of the female breast images (for the National Cancer Institute 2011; Terese Winslow, US Government, has certain rights).
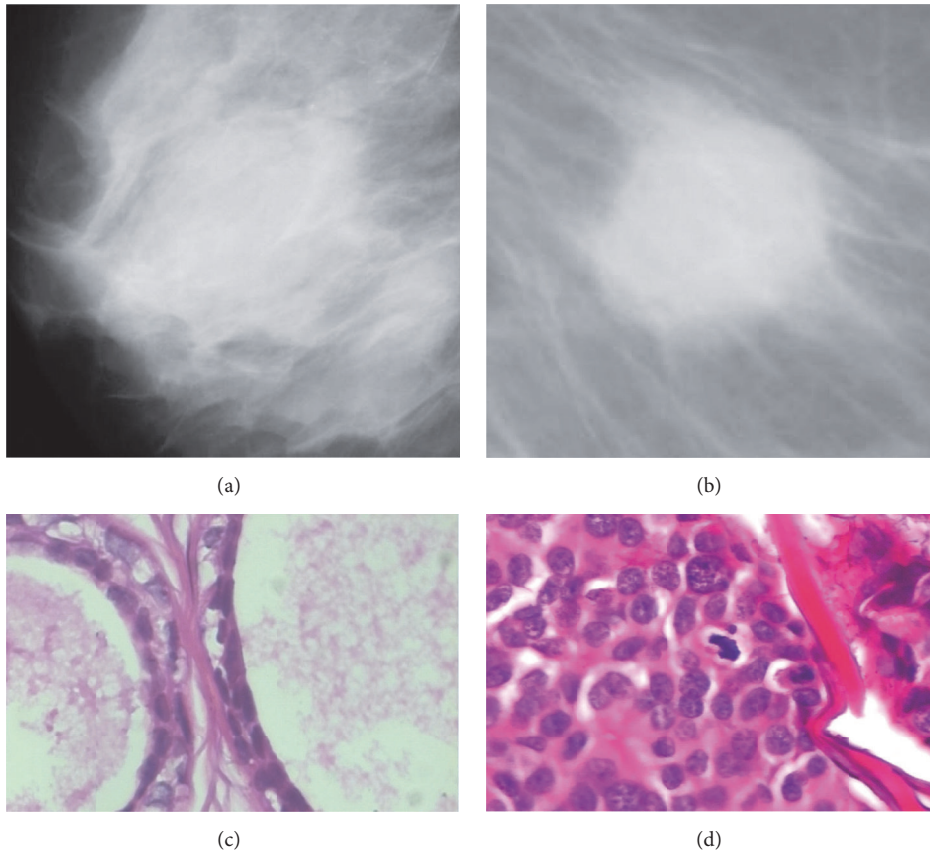

(a)

(b)

(c)

(d)

Figure 4: (a, b) show mammogram benign and malignant images (examples of noninvasive image) and (c, d) show histopathological benign and malignant images (examples of invasive image).

Table 1: Available breast image database for biomedical investigation.

| Database	Number of images	Database size (GB)	Image capture technique	Image type	Total patients
MIAS	322	2.3	Mammogram		161
DDSM			Mammogram		2620
CBIS-DDSm	4067	70.5	MG	DICOM	237
ISPY1	386,528	76.2	MR, SEG		237
Breast-MRI-NACT-Pilot	99,058	19.5	MRI		64
QIN-Breast	100835	11.286	PET/CT, MR	DICOM	67
Mouse-Mammary	23487	8.6	MRI	DICOM	32
TCGA-BRCA	230167	88.1	MR, MG	DICOM	139
QIN Breast DCE-MRI	76328	15.8	CT	DICOM	10
BREAST-DIAGNOSIS	105050	60.8	MRI/PET/CT	DICOM	88
RIDER Breast MRI	1500	.401	MR	DICOM	5
BCDR			Mammogram		1734
TCGA-BRCA		53.92 (TB)	Histopathology		1098
BreakHis	7909		Histopathology		82
Inbreast	419		Mammogram		115
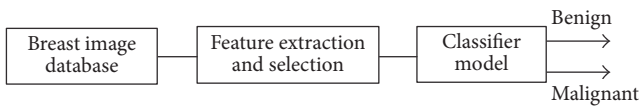
Figure 5: A very basic breast image classification model.

outcome, researchers always base their investigation on some well-established image database. Various organizations have introduced sets of images databases which are available to researchers for further investigation. Table 1 gives a few of the available image databases, with some specifications.

The image formats of the different databases are different. Few of the images contained images in JPEG format and few databases contained DICOM-format data. Here the MIAS, DDSM, and Inbreast databases contain mammogram images. According to the Springer (http://www.springer.com), Elsevier (https://www.elsevier.com), and IEEE (http://www.ieeexplore.ieee.org) web sites, researchers have mostly utilized the MIAS and DDSM databases for the breast image classification research. The number of conference papers published for the DDSM and MIAS databases is 110 and 168, respectively, with 82 journal papers published on DDSM databases and 136 journal papers published using the MIAS database. We have verified these statistics on both Scopus (https://www.scopus.com) and the Web of Science database (http://www.webofknowledge.com). Figure 6 shows the number of published breast image classification papers based on the MIAS and DDSM database from the years 2000 to 2017.

Histopathological images provide valuable information and are being intensively investigated by doctors for finding the current situation of the patient. The TCGA-BRCA and BreakHis databases contain histopathological images. Research has been performed in a few experiments on this database too. Among these two databases, BreakHis is the most recent histopathological image database, containing a
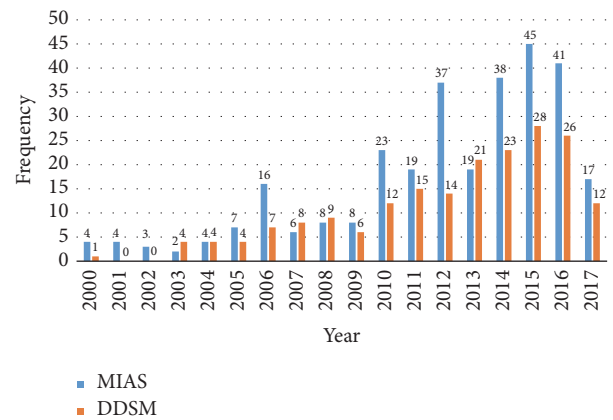

Figure 6: Number of papers published based on MIAS and DDSM databases.

total of 7909 images which have been collected from 82 patients [6]. So far around twenty research papers have been published based on this database.

2.2. Feature Extraction and Selection. An important step of the image classification is extracting the features from the images. In the conventional image classification task, features are crafted locally using some specific rules and criteria. However, the-state-of-the-art Convolutional Neural Network (CNN) techniques generally extract the features globally using kernels and these Global Features have been used for image classification. Among the local features, texture, detector, and statistical are being accepted as important features for breast image classification. Texture features actually represent the low-level feature information of an image, which provides more detailed information of an image that might be possible from histogram information alone. More specifically, texture features provide the structural and dimensional information of the color as well as the intensity

TABLE 2: Feature descriptor.

| Feature category	Feature description
Texture	Haralick texture features [7] (1) Angular Second Moment (ASM), (2) Contrast, (3) correlation, (4) Sum of Squares of Variances (SSoV), (5) Inverse of Difference (IoD), (6) Sum of Average (SoA), (7) Sum of Variances (SoV), (8) Sum of Entropy (SoE), (9) Entropy, (10) Difference of Variance (DoV), (11) Difference of Entropy (DoE), (12) Gray-Level Concurrence Matrix (GLCM). Tamura features [8] (1) Coarseness, (2) Contrast, (3) directionality, (4) line-likeness, (5) roughness, (6) regularity. Global texture descriptor (1) Fractal dimension (FD), (2) Coarseness, (3) Entropy, (4) Spatial Gray-Level Statistics (SGLS), (5) Circular Moran Autocorrelation Function (CMAF).
Detector	Single scale detector (1) Moravec's Detector (MD) [9], (2) Harris Detector (HD) [10], (3) Smallest Univalue Segment Assimilating Nucleus (SUSAN) [11], (4) Features from Accelerated Segment Test (FAST) [12, 13], (5) Hessian Blob Detector (HBD) [14, 15]. Multiscale detector [8] (1) Laplacian of Gaussian (LoG) [9, 16], (2) Difference of Gaussian (DoG) Contrast [17] (3) Harris Laplace (HL), (4) Hessian Laplace (HeL), (5) Gabor-Wavelet Detector (GWD) [18].
Strutural	(1) Area, (2) bounding box, (3) centroid, (4) Convex Hull (CH), (5) eccentricity, (6) Convex Image (CI), (7) compactness, (8) Aspect Ratio (AR), (9) moments, (10) extent, (11) extrema, (12) Major Axis Length (MaAL), (13) Minor Axis Length (MiAL), (14) Maximum Intensity (MaI), (15) Minimum Intensity (MiI), (16) Mean Intensity (MI), (17) orientation, (18) solidity.
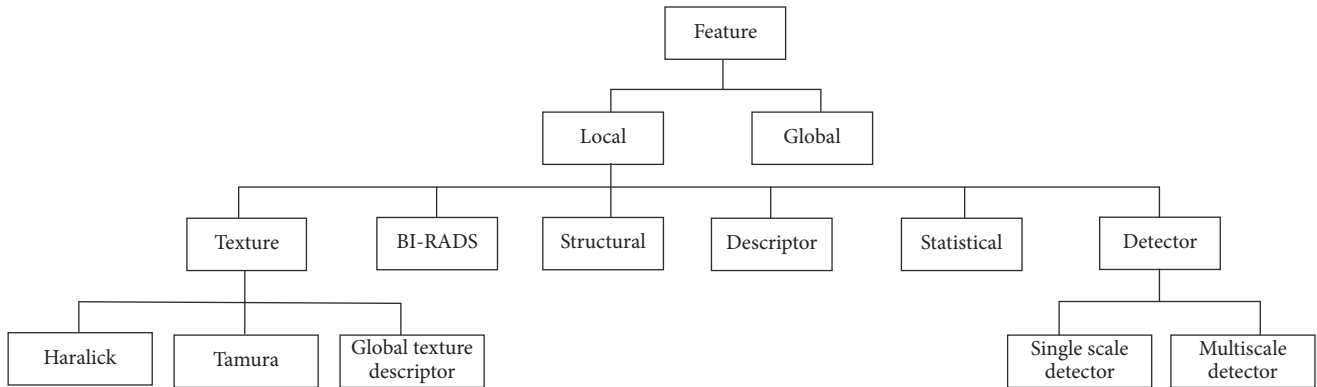
FIGURE 7: Classification of features for breast image classification.

of the image. Breast Imaging-Reporting and Data System (BI-RADS) is a mammography image assessment technique, containing 6 categories normally assigned by the radiologist. Feature detector actually provides information whether the particular feature is available in the image or not. Structural features provide information about the features structure and orientation such as the area, Convex Hull, and centroid. This kind of information gives more detailed information about the features. In a cancer image, it can provide the area of the nucleus or the centroid of the mass. Mean, Median, and Standard Deviation always provide some important information on the dataset and their distribution. This kind of features has been categorized as statistical features. The total hierarchy of the image feature extraction is resented in Figure 7. Tables 2 and 3 further summarize the local features in detail.

Features which are extracted for classification do not always carry the same importance. Some features may even contribute to degrading the classifier performance. Prioritization of the feature set can reduce the classifier model complexity and so it can reduce the computational time. Feature set selection and prioritization can be classified into three broad categories:

(i) Filter: the filter method selects features without evaluating any classifier algorithm.

(ii) Wrapper: the wrapper method selects the feature set based on the evaluation performance of a particular classifier.

(iii) Embedded: the embedded method takes advantage of the filter and wrapper methods for classifier construction.

TABLE 3: Feature descriptor.

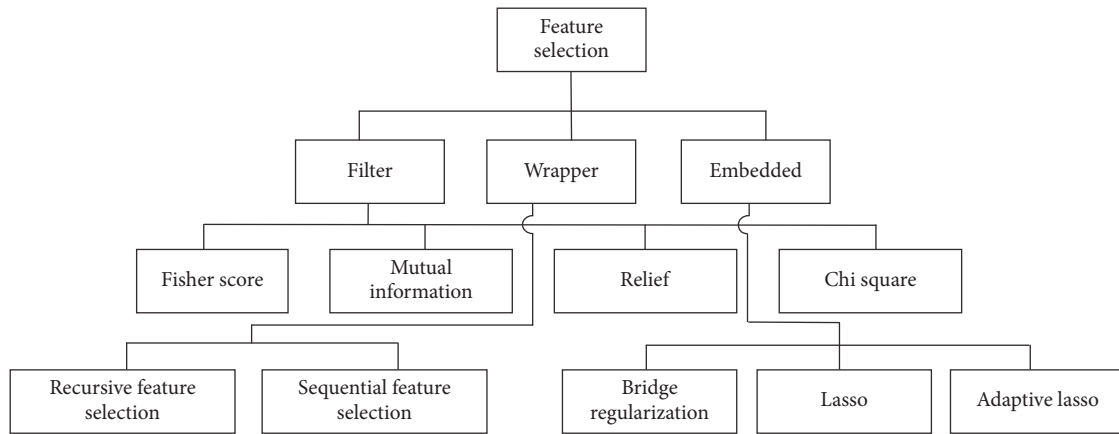| Feature category	Feature description
Statistical	(1) Mean, (2) Median, (3) Standard Deviation, (4) Skewness, (5) Kurtosis, (6) Range
Descriptor	(1) Scale Invariant Feature Transform (SIFT) [17, 19], (2) Gradient Location-Orientation Histogram (GLOH) [20], (3) Speeded-Up Robust Features Descriptor (SURF) [21–23], (4) Local Binary Pattern (LBP), [24–27], (5) Binary Robust Independent Elementary Features (BRIEF) [28], (6) Weber Local Descriptor (WLD) [29, 30] (7) Back Ground Local Binary Pattern (BGLBP) [31], (8) Center-Symmetric Local Binary Pattern (CS-LBP), [32] (9) Second-Order Center-Symmetric Local Derivative Pattern (CS-LBP) [33], (10) Center-Symmetric Scale Invariant Local Ternary Patterns (CS-SILTP) [34], (11) Extended LBP or Circular LBP (E-LBP) [35], (12) Opponent Color Local Binary Pattern (OC-LBP), [36] (13) Original LBP(O-LBP) [25], (14) Spatial Extended Center-Symmetric Local Binary Pattern (SCS-LBP) [37], (15) Scale Invariant Local Ternary Pattern (SI-LTP) [38], (16) Variance-Based LBP (VAR-LBP) [24], (17) eXtended Center-Symmetric Local Binary Pattern (XCS-LBP), (18) Average Local Binary Pattern (ALBP), (19) Block Based Local Binary Pattern (BBLBP) [39],
BI-RADS [40]	(1) Margin Integrality (MarI), (2) Margin Ambiguity (MarA), (3) Echo Pattern Posterior Feature (EPPF), (4) Calcification in Mass (CM), (5) Architectural Distortion (AD), (6) Edema, (7) Eymph Nodes Axillary (ENA) (8) Ducts Changes (DC), (9) Skin Thickening (ST), (10) Postsurgical Fluid Collection (PSFC), (11) Skin Retraction (SR1), (12) Fat Necrosis (FN), (13) Lump Nodes Intramammary (LNI).

FIGURE 8: A summary of feature selection method.

Figure 8 shows a generalized feature selection method where we have further classified the filter method into Fisher Score, Mutual Information, Relief, and chi square methods. The embedded method has been classified into Bridge Regularization, Lasso, and Adaptive Lasso methods, while the wrapper method has been classified to recursive feature selection and sequential feature selection method.

2.3. Classifier Model. Based on the learning point of view, breast image classification techniques can be categorized into the following three classes [41]:

  (i) Supervised

  (ii) Unsupervised

  (iii) Semisupervised.

These three classes can be split into Deep Neural Network (DNN) and conventional classifier (without DNN) and to some further classes as in Table 4.

2.4. Performance Measuring Parameter. A Confusion Matrix is a two-dimensional table which is used to a give a visual

FIGURE 9: Confusion Matrix.

perception of classification experiments [54]. The (i, j)th position of the confusion table indicates the number of times that the ith object is classified as the jth object. The diagonal of this matrix indicates the number of times the objects are correctly classified. Figure 9 shows a graphical representation of a Confusion Matrix for the binary classification case.

TABLE 4: A simplified hierarchy of classification.

Learning technique		Algorithm	
Supervised	Conventional	(a) Logic based	(1) ID3, (2) C4.5, (3) bagging, (4) random trees, (5) Random Forest, (6) boosting, (7) advanced boosting, (8) Extreme Boosting (XGBoosting).
		(b) Bayesian	(1) Naive Bayes
			(2) Bayesian Network
		(c) Conventional Neural Network	
		(d) Support Vector Machine	
	DNN-based	(a) Convolutional Neural Network (CNN),	
		(b) Deep Belief Network (DBN),	
		(c) Generative Adversarial Network (GAN).	
Unsupervised	Conventional	(a) k-Means Clustering	
		(b) Self-Organizing Map (SOP)	
		(c) Fuzzy C-Means Clustering (FCM)	
	DNN-based	(a) Deep Belief Network (DBN)	
Semisupervised	Conventional	(a) Self-training	
		(b) Graph Based	
		(c) S3V3	
		(d) Multiview	
		(e) Generative model	

Among the different classification performance properties, this matrix will provide following parameters:

(i) Recall is defined as Recall = TP/(TP + FN).

(ii) Precision is defined as Precision = TP/(TP + FP).

(iii) Specificity is defined as Specificity = TN/(TN + FP).

(iv) Accuracy is defined as ACC = (TP + TN)/(TP + TN + FP + FN).

(v) F-1 score is defined as F_1 = (2 × Recall)/(2 × Recall + FP + FN).

(vi) Matthew Correlation Coefficient (MCC): MCC is a performance parameter of a binary classifier, in the range {−1 to +1}. If the MCC values trend more towards +1, the classifier gives a more accurate classifier and the opposite condition will occur if the value of the MCC trend towards the −1. MCC can be defined as

$$
\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FP)}}. \tag{1}
$$

3. Performance of Different Classifier Model on Breast Images Dataset

Based on Supervised, Semisupervised, and Unsupervised methods different research groups have been performed classification operation on different image database. In this section we have summarized few of the works of breast image classification.

3.1. Performance Based on Supervised Learning. In supervised learning, a general hypothesis is established based on externally supplied instances to produce future prediction. For the supervised classification task, features are extracted or automatically crafted from the available dataset and each sample is mapped to a dedicated class. With the help of the features and their levels a hypothesis is created. Based on the hypothesis unknown data are classified [55].

Figure 10 represents an overall supervised classifier architecture. In general, the whole dataset is split into training and testing parts. To validate the data, some time data are also split into a validation part as well. After the data splitting the most important part is to find out the appropriate features to classify the data with the utmost Accuracy. Finding the features can be classified into two categories, locally and globally crafted. Locally crafted means that this method requires a hand-held exercise to find out the features, whereas globally crafted means that a kernel method has been introduced for the feature extraction. Handcrafted features can be prioritized, whereas Global Feature selection does not have this luxury.

3.1.1. Conventional Neural Network. The Neural Network (NN) concept comes from the working principle of the human brain. A biological neuron consists of the following four parts:
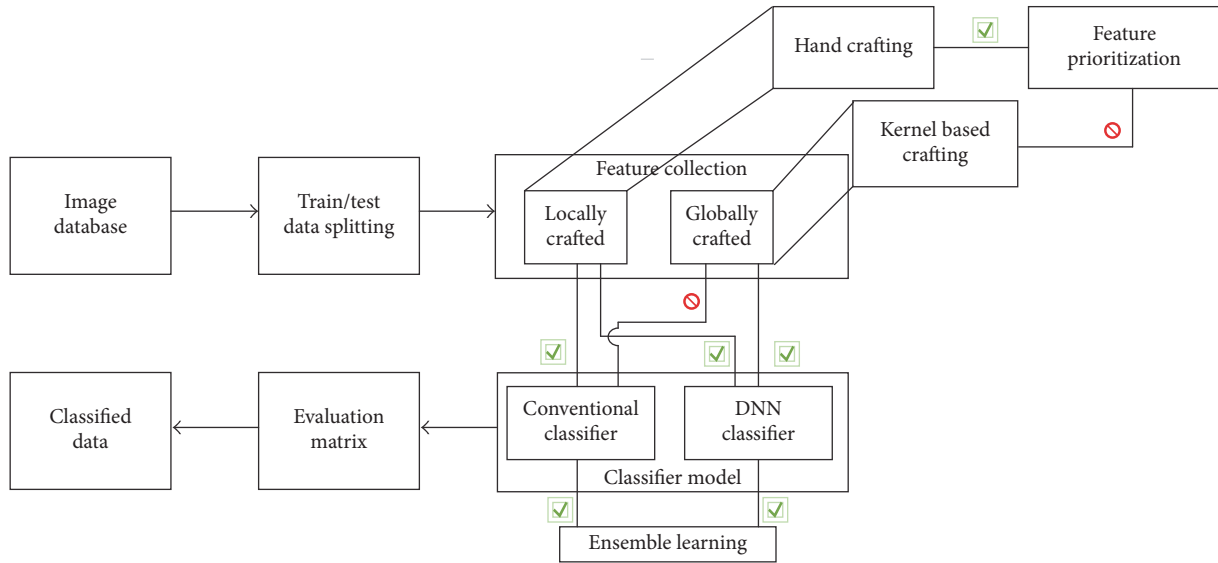
(i) Dendrites

(ii) Nuclease

(iii) Cell body

(iv) Axon.
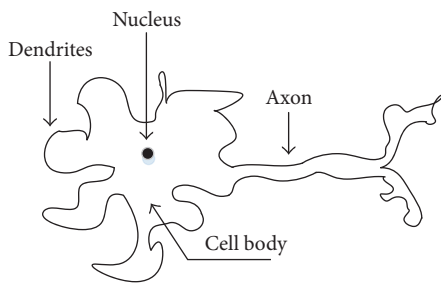
FIGURE 10: A generalized supervised classifier model.

FIGURE 11: A model of a biological neuron.

FIGURE 12: Working principle of a simple Neural Network technique.

Dendrites collect signals and axons carry the signal to the next dendrite after processing by the cell body as shown in Figure 11. Using the neuron working principle, the perceptron model was proposed by Rosenblatt in 1957 [56]. A single-layer perceptron linearly combines the input signal and gives a decision based on a threshold function. Based on the working principle and with some advanced mechanism and engineering, NN methods have established a strong footprint in many problem-solving issues. Figure 12 shows the basic working principle of NN techniques.

In the NN model the input data $\mathbf{X} = \{x_0, x_1, \ldots, x_N\}$ is first multiplied by the weight data $\mathbf{W} = \{w_0, w_1, \ldots, w_N\}$ and then the output is calculated using

$$\mathbf{Y} = \mathbf{g}\left(\sum\right) \quad \text{where } \sum = \mathbf{W} \cdot \mathbf{X}. \qquad (2)$$

Function $\mathbf{g}$ is known as the activation function. This function can be any threshold value or Sigmoid or hyperbolic and so forth. In the early stages, feed-forward Neural Network techniques were introduced [57]; lately the backpropagation method has been invented to utilize the error information to improve the system performance [58, 59].

The history of breast image classification by NN is a long one. To the best of my knowledge a lot of the pioneer work

was performed by Dawson et al. in 1991 [60]. Since then, NN has been utilized as one of the strong tools for breast image classification. We have summarized some of the work related to NN and breast image classification in Tables 5, 6, and 7.

3.1.2. Deep Neural Network. Deep Neural Network (DNN) is a state-of-the-art concept where conventional NN techniques have been utilized with advanced engineering. It is found that conventional NNs have difficulties in solving complex problems, whereas DNNs solve them with utmost Precision. However DNNs suffer from more time and computational complexity than the conventional NN.

(i) Convolutional Neural Network (CNN)

(ii) Deep Belief Network (DBN)

(iii) Generative Adverbial Network (GAN)

(iv) Recurrent Neural Network (RNN)

Convolutional Neural Network. A CNN model is the combination of a few intermediate mathematical structures. This intermediate mathematical structure creates or helps to create different layers:

(i) Convolutional Layer. Among all the other layers, the convolutional layer is considered as the most important part for a CNN model and can be considered as the backbone of

TABLE 5: Neural Network for breast image classification.

Reference	Descriptor	Image type	Number of images	Key findings
Rajakeerthana et al. [42]	(1) GLCM, GLDM, SRDM, NGLCM, GLRM	Mammogram	322	(1) The classifier achieved 99.20% Accuracy.
Lessa and Marengoni [43]	(1) Mean, Median, Standard Deviation, Skewness, Kurtosis, Entropy, Range	Thermographic	94	(1) Achieved Sensitivity, Specificity, and Accuracy are 87.00%, 83.00%, and 85.00%, respectively.
Wan et al. [44]	(1) ALBP (2) BBLBP	OCM	46	(1) Achieved Sensitivity and Specificity are 100% and 85.20. respectively. (2) ROC value obtained 0.959.
Chen et al. [40]	(1) 19 BI-RADS features have been used	Ultrasound	238	(1) Chi squared method has been utilized for the feature selection. (2) Achieved Accuracy, Sensitivity, and Specificity are 96.10%, 96.70%, and 95.70%, respectively.
de Lima et al. [45]	(1) Total 416 features have been used	Mammogram	355	(1) Multiresolution wavelet and Zernike moment have been utilized for the feature extraction.
Abirami et al. [46]	(1) 12 statistical measures such as Mean, Median, and Max have been utilized as the features	Mammogram	322	(1) Wavelet transform has been utilized for the feature extraction. (2) The achieved Accuracy, Sensitivity, and Specificity are 95.50%, 95.00%, and 96.00%, respectively.
El Atlas et al. [47]	(1) 13 morphological features have been utilized	Mammogram	410	(1) Firstly the edge information has been utilized for the mass segmentation and then the morphological features were extracted. (2) Achieved best Accuracy is 97.5%.

TABLE 6: Neural Network for breast image classification.

Reference	Descriptor	Image type	Number of images	Key findings
Alharbi et al. [48]	(1) 49 features have been utilized.	Mammogram	1100	(1) Five feature selection methods: Fisher score, Minimum Redundancy-Maximum Relevance, Relief-f, Sequential Forward Feature Selection, and Genetic Algorithm have been used. (2) Achieved Accuracy, Sensitivity, and specificity are 94.20%, 98.36%, and 99.27%, respectively
Peng et al. [49]	(1) Haralick and Tamura features have been utilized	Mammogram	322	(1) Feature reduction has been performed by Rough-Set theory and selected 5 prioritized features. (2) The best Accuracy, Sensitivity, and Specificity achieved were 96.00%, 98.60%, and 89.30%
Jalalian et al. [50]	(1) GLCM (2) Compactness	Mammogram		(1) The obtained classifier Accuracy, Sensitivity, and Specificity are 95.20%, 92.40%, and 98.00%, respectively.
Li et al. [51]	(1) Four feature vectors have been calculated	Mammogram	322	(1) 2D contour of breast mass in mammography has been converted into 1D signature. (2) NN techniques achieved Accuracy is 99.60% when RMS slope is utilized.
Chen et al. [52]	(1) Autocorrelation features	Ultrasound	242	(1) The overall achieved Accuracy, Sensitivity, and Specificity are 95.00%, 98.00%, and 93%, respectively.
Chen et al. [53]	(1) Autocorrelation features	Ultrasound	1020	(1) The obtained ROC area is 0.9840 ± 0.0072.

TABLE 7: Neural Network for breast image classification.

Reference	Descriptor	Image type	Number of images	Key findings
Chen et al. [61]	(1) Variance Contrast of Wavelet Coefficient (2) Autocorrelation of Wavelet Coefficient	Ultrasound	242	(1) The achieved ROC curve 0.9396 ± 0.0183
Silva et al. [62]	(1) 22 different morphological features such as convexity and lobulation have been utilized	Ultrasound	—	(1) The best obtained Accuracy and ROC curve are 96.98% and 0.98, respectively
Saritas [63]	(1) Age of patient, (2) mass shape, (3) mass border, (4) Mass density, (5) BIRADS	Mammogram	—	(1) Disease prediction rate is 90.5% (2) Neural Network utilized 5 neurons in input layers and one hidden layer.
López-Meléndez et al. [64]	(1) Area, perimeter, etc. have been utilized	Mammogram	322	(1) The achieved Sensitivity and Specificity are 96.29% and 99.00%, respectively.

the model. A kernel of size $m \times n$ is scanned through the input data for the convolutional operation which ensures the local connectivity and weight sharing property.

(ii) Stride and Padding. In the convolutional operation, a filter scans through the input matrices. In each step how much position a kernel filter moves through the matrix is known as the stride. By default stride keeps to 1. With inappropriate selection of the stride the model can lose the border information. To overcome this issue the model utilizes extra rows and columns at the end of the matrices, and these added rows and columns contain all 0s. This adding of extra rows and columns which contain only zero value is known as zero padding.

(iii) Nonlinear Operation. The output of each of the kernel operations is passed through a rectifier function such as Rectified Linear Unit (ReLU), Leaky-ReLU, TanH, and Sigmoid. The Sigmoid function can be defined as

$$\sigma(x) = \frac{1}{(1 + \exp^{-x})} \tag{3}$$

and the tanh function can be defined as

$$\tanh(x) = \frac{(\exp^x - \exp^{-x})}{(\exp^x + \exp^{-x})}. \tag{4}$$

However the most effective rectifier is ReLU. The ReLU method converts all the information into zero if it is less than or equal to zero and passes all the other data as is shown in Figure 13

$$\sigma(x) = \max(0, x). \tag{5}$$

Another important nonlinear function is Leaky-RelU

$$\text{Leaky-ReLU}(x) = \sigma(x) + \alpha \min(0, x), \tag{6}$$

where α is predetermined parameter which can be varied to give a better model.
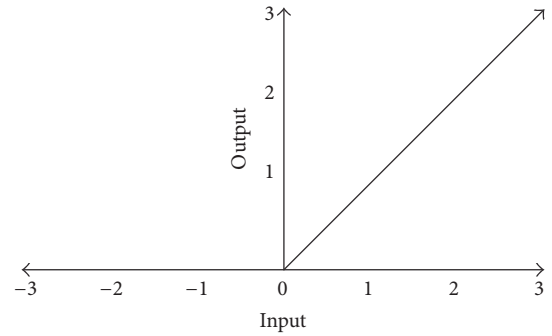
FIGURE 13: ReLU Operation.

(iv) Subsampling. Subsampling is the procedure of reducing the dimensionality of each of the feature maps of a particular layer; this operation is also known as a pooling operation. Actually it reduces the amount of feature information from the overall data. By doing so, it reduces the overall computational complexity of the model. To do this $s \times s$ patch units are utilized. The two most popular pooling methods are

(a) Max-Pooling

(b) Average Pooling.

In Max-Pooling, only the maximum values within a particular kernel size are selected for further calculation. Consider an example of a 16×16 image as shown in Figure 14. A 2 by 2 kernel is applied to the whole image, 4 blocks in total, and produces a 4×4 output image. For each block of four values, we have selected the maximum. For instance, from blocks one, two, three, and four, maximum values 4, 40, 13, and 8 are selected, respectively, as they are the maximum in that block. For the Average Pooling operation, each kernel gives the output as average.

(v) Dropout. Regularization of the weight can reduce the outfitting problem. Randomly removing some neurons can

FIGURE 14: Max-Pooling and Average Pooling.

FIGURE 15: Work-flow of a Convolutional Neural Network.

regularize the overfilling problem. The technique of randomly removing neurons from the network is known as dropout.

(vi) Soft-Max Layer. This layer contains normalized exponential functions to calculate the loss function for the data classification.

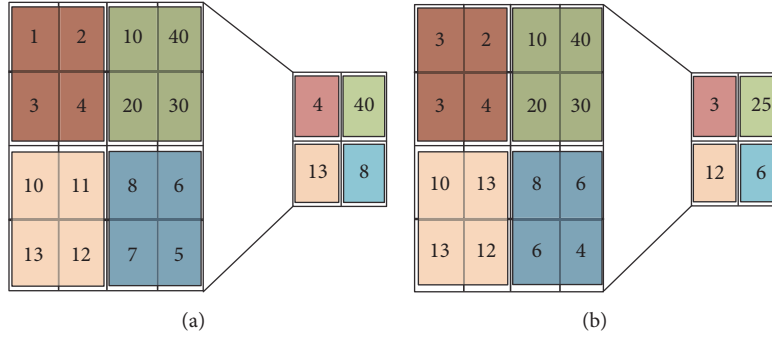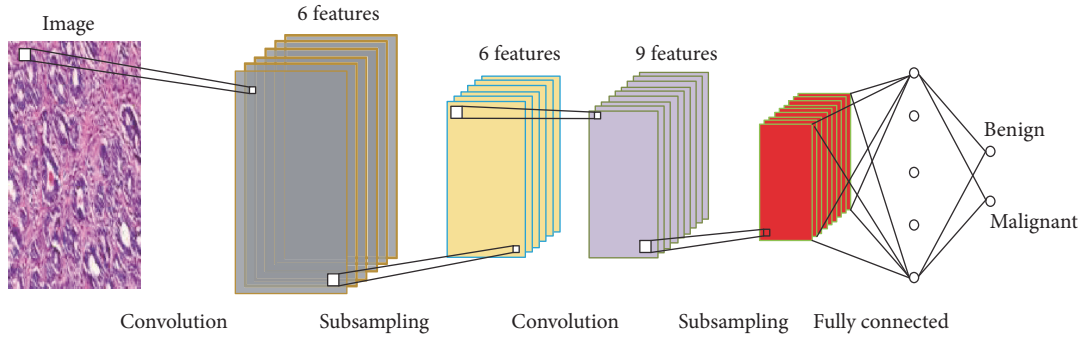Figure 15 shows a generalized CNN model for the image classification. All the neurons of the most immediate layer of a fully connected layer are completely connected with the fully connected layer, like a conventional Neural Network. Let f_j^{l-1} represent the jth feature map at the layer $l-1$. The jth feature map at the layer l can be represented as

$$f_j^l = \sigma \left(\sum_{i=1}^{N^{l-l}} f_i^{l-1} * k_{i,j} + b_j^l \right), \tag{7}$$

where N^{l-l} represents the number of feature maps at the $l-1$th layer, $k_{i,j}$ represents the kernel function, and b_j^l represents the bias at l, where σ performs a nonlinear function operation. The layer before the Soft-Max Layer can be represented as

$$h_p^{\text{end}} = w^{\text{end}} * h_p^{\text{end}-1} + b^{\text{end}}. \tag{8}$$

As we are working on a binary classification, the Soft-Max regression normalized output can be represented as

$$\overline{y}_p = \frac{\exp\left(h_p^{\text{end}}\right)}{\sum_{p=1}^2 \exp\left(h_p^{\text{end}}\right)}. \tag{9}$$

Let $p = 1$ represent Benign class and $p = 2$ represent the Malignant class. The cross-entropy loss of the above function can be calculated as

$$L_p = -\ln\left(\overline{y}_p\right). \tag{10}$$

Whichever group experiences a large loss value, the model will consider the other group as predicted class.

A difficult part of working on DNN is that it requires a specialized software package for the data analysis. Few research groups have been working on how effectively data can be analyzed by DNN from different perspectives and the demand. Table 8 summarizes some of the software which is available for DNN analysis.

The history of the CNN and its use for biomedical image analysis is a long one. Fukushima first introduced a CNN named "necognitron" which has the ability to recognize stimulus patterns with a few shifting variances [113]. To the best of our knowledge, Wu et al. first classified a set of mammogram images into malignant and benign classes using a CNN model [78]. In their proposed model they only utilized one hidden layer. After that, in 1996 Sahiner et al. utilized CNN model to classify mass and normal breast tissue and achieved ROC scores of 0.87 [79]. In 2002, Lo et al. utilized a Multiple Circular Path CNN (MCPCNN) for tumor identification from mammogram images and obtained ROC scores of around 0.89. After an absence of investigation of the CNN model, this model regained its momentum after the work of Krizhevsky et al. [114]. Their proposed model is known as AlexNet. After this work a revolutionary change

TABLE 8: Available software for deep learning analysis.

Software	Interface and backend	Provider
Caffe [65, 66]	Python, MATLAB, C++	Berkeley Vision and Learning Centre, University of California, Berkeley
Torch [67]	C, LuaJIT	
MatConvNet [68, 69]	MATLAB, C	Visual Geometry Group, Department of Engineering, University of Oxford
Theano [70, 71]	Python	Montreal Institute for Learning Algorithms University of Montreal
TensorFlows [72]	C++, Python	Google
CNTK [73]	C++	Microsoft
Keras [74]	Theano, Tensor Flow	MIT
dl4j [75]	Java	Skymind Engineering
DeeBNET [76, 77]	MATLAB	Information Technology Department, Amirkabir University of Technology

has been achieved in the image classification and analysis field. As an advanced engineering of the AlexNet, the paper titled "Going Deeper with Convolutions" by Szegedy [115] introduced the GoogleNet model. This model contains a much deeper network than AlexNet. Sequentially ResNet [116], Inception [117], Inception-v4, Inception-ResNet [118], and a few other models have recently been introduced.

Later, directly or with some advanced modification, these DNN models have been adapted for biomedical image analysis. In 2015, Fonseca et al. [81] classified breast density using CNN techniques. CNN requires a sufficient amount of data to train the system. It is always very difficult to find a sufficient amount of medical data for training a CNN model. A pretrained CNN model with some fine tuning can be used rather than create a model from scratch [119]. The authors of [119] did not perform their experiments on a breast cancer image dataset; however they have performed their experiments on three different medical datasets with layer-wise training and claimed that "retrained CNN along with adequate training can provide better or at least the same amount of performance."

The Deep Belief Network (DBN) is another branch of the Deep Neural Network, which mainly consists of Restricted Boltzmann Machine (RBM) techniques. The DBN method was first utilized for supervised image classification by Liu et al. [120]. After that, Abdel-Zaher and Eldeib utilized the DBN method for breast image classification [121]. This field is still not fully explored for breast image classification yet. Zhang et al. utilized both RBM and Point-Wise Gated RBM (PRBM) for shear-wave electrography image classification where the dataset contains 227 images [97]. Their achieved classification Accuracy, Sensitivity, and Specificity are 93.40%, 88.60%, and 97.10%, respectively. Tables 9, 10, and 11 have summarized the most recent work for breast image classification along with some pioneer work on CNN.

3.1.3. Logic Based Algorithm. A Logic Based algorithm is a very popular and effective classification method which follows the tree structure principle and logical argument as shown in Figure 16. This algorithm classifies instances based on the feature's values. Along with other criteria, a decision-tree based algorithm contains the following features:

(i) Root node: a root node contains no incoming node, and it may or may not contain any outgoing edge

(ii) Splitting: splitting is the process of subdividing a set of cases into a particular group. Normally the following criteria are maintained for the splitting:

    (a) information gain,

    (b) Gini index,

    (c) chi squared

(iii) Decision node

(iv) Leaf/terminal node: this kind of node has exactly one incoming edge and no outgoing edge. The tree always terminates here with a decision

(v) Pruning: pruning is a process of removing subtrees from the tree. Pruning performs to reduce the overfitting problem. Two kinds of pruning techniques are available:

    (a) prepruning,

    (b) postpruning.

Among all the tree based algorithms, Iterative Dichotomiser 3 (ID3) can be considered as a pioneer, proposed by Quinlan [149]. The problem of the ID3 algorithm is to find the optimal solution which is very much prone towards overfitting. To overcome the limitation of the ID3 algorithm the C4.5 algorithm has been introduced by Quinlan [150], where a pruning method has been introduced to control the overfitting problem. Pritom et al. [151] classified the Wisconsin breast dataset where they utilized 35 features. They have obtained 76.30% Accuracy, 75.10% False Positive Rate, and ROC score 0.745 when they ranked the features. Without ranking the features they obtained 73.70% Accuracy, 50.70% False Positive Rate, and ROC score value 52.80. Asri et al. [152] utilized the C4.5 algorithm for the Wisconsin

TABLE 9: Convolutional Neural Network.

Reference	Descriptor	Image type	Number of images	Key findings
Wu et al. [78]	(1) Global Features	Mammogram	40	(1) Achieved Sensitivity 75.00% and Specificity 75.00%.
Sahiner et al. [79]	(1) Global Features	Mammogram	168	(1) The achieved ROC score is 0.87.
Lo et al. [80]	(1) Density, size, Shape, Margin	Mammogram	144	(1) The achieved ROC curve is 0.89.
Fonseca et al. [81]	(1) Global Features	Mammogram	—	(1) Breast density classification has been performed utilizing HT-L3 convolution. (2) Average achieved obtained Kappa value is 0.58.
Arevalo et al. [82]	(1) Global Features	Mammogram	736	(1) The achieved ROC curve is 0.826.
Su et al. [83]	(1) Global Features	Mammogram	92	(1) Fast Scanning CNN (fCNN) method has been utilized to reduce the information loss. (2) The average Precision, Recall, and $F1$ score are 91.00%, 82.00%, and 0.85, respectively.
Sharma and Preet [84]	(1) GLCM, GLDM Geometrical	Mammogram	40	(1) The best Accuracy achieved is 75.23% and 72.34%, respectively, for fatty and dense tissue classification.
Spanhol et al. [6]	(1) Global Features	Histopathology	7909	(1) The best Accuracy achieved 89 ± 6.6%.
Rezaeilouyeh et al. [85]	(1) Local and Global Features	Histopathology	—	(1) Shearlet transform has been utilized for extracting local features. (2) When they utilize RGB image along with magnitude of Shearlet transform together, the Achieved Sensitivity, Specificity, and Accuracy were 84.00 ± 1.00%, 91.00 ± 2.00%, and 84.00 ± 4.00%; when they utilize RGB image along with both the phase and magnitude of Shearlet transform together, the achieved Sensitivity, Specificity, and Accuracy were 89.00 ± 1.00%, 94.00 ± 1.00%, and 88.00 ± 5.00%.

FIGURE 16: A general structure of a tree.

database classification where they utilized 11 features and obtained 91.13% Accuracy.

Logic Based algorithms allow us to produce more than one tree and combine the decisions of those trees for an advanced result; this mechanism is known as an ensemble method. An ensemble method combines more than one classifier hypothesis together and produces more reliable results through a voting concept. Boosting and bagging are two well-known ensemble methods. Both boosting and bagging aggregate the trees. The difference is in bagging successive trees do not depend on the predecessor trees, where in the boosting method successive trees depend on the

TABLE 10: Convolutional Neural Network.

Reference	Descriptor	Image type	Number of images	Key findings
Albayrak and Bilgin [86]	(1) Global Features	Histopathology	100	(1) Cluster-based segmentation has been performed to find out the cellular structure. (2) Blob analysis has been performed on the segmented images. (3) To reduce the high dimensionality, Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) methods have been utilized. (4) Before the dimensionality reduction the Precision, Recall, and F-score values were 97.20%, 66.00%, and 0.78%, respectively, but when the dimensionality reduction method was utilized the Precision, Recall, and F-score values were 100.00%, 94.00%, and 0.96%, respectively (5) The best average Accuracy is 73.00% (without dimensionality reduction) and 96.8% (with dimensionality reduction).
Jiao et al. [87]	(1) Global and Local Features.	Mammogram	—	(1) They performed their experiments on the DDSM database. (2) Total required parameter is 5.8×10^7 and time for the per image processing is 1.10 ms. (3) The best classification achieved is 96.70%; however they show that when they utilize the VGG model the Accuracy was 97.00% which is slightly better than their model. However in terms of memory size and time per image processing their model gives better performance than the VGG model.
Zejmo et al. [88]	(1) Global Features	Cytology	40	(1) GoogleNet and AlexNet models have been utilized. (2) The best Accuracy obtained when they utilized GoogleNet model was 83.00%.

information gathered from the predecessor trees. Gradient boosting is a very popular method for data classification [153, 154]; however a state-of-the-art boosting algorithm such as "Extreme Gradient Boosting" (XGBoosting) is a very effective method for data classification [155]. Interestingly, there has not been a single paper published for breast image classification using the XGBoost algorithm. Along with the boosting method, different bagging methods are available; among them Random Forest (RF) is very popular where a large number of uncorrelated trees are aggregated together for a better prediction. Tables 12 and 13 summarize a set of papers where a Logic Based algorithm has been used for image classification.

3.1.4. Support Vector Machine (SVM). SVM were proposed by VC (Vepnick-Cherovorenkis). This technique does not require any prior distribution knowledge for the data classification task like Bayesian classification technique. In many practical situations, the distribution of the features is not available. In such cases, SVM can be used to classify the available data into the different classes.

Consider the set of two-dimensional data plotted in Figure 17. The symbol "∘" represents those data which belong to Class-1 and "□" represents data which belong to Class-2. A hyperplane (P) has been drawn which classifies the data into two classes. Interestingly, there will be "n" hyperplanes available which can separate the data.

Let $\mathbf{X} = \{\mathbf{X}_i\}$, where $\{\mathbf{X}_i \in \mathscr{R}^n\}$ ($i = \{1, 2, 3, \ldots, l\}$) is to be classified into two classes $\omega \in \{\omega^1, \omega^2\}$. Suppose that the classes $\{\omega^1\}$ and $\{\omega^2\}$ are recognized as "+1" and "−1". Classification of this data can be written

$$\mathscr{C} = \{(\mathbf{X}_1, \omega_1), (\mathbf{X}_2, \omega_2), (\mathbf{X}_3, \omega_3), \ldots, (\mathbf{X}_n, \omega_n)\}. \quad (11)$$

During the learning stage, the SVM finds parameters $\mathbf{W}_i = [W_i^1, W_i^2, \ldots, W_i^n]^T$ and b to produce a decision function $d(\mathbf{X}_i, \mathbf{W}_i, b)$:

$$d(\mathbf{X}_i, \mathbf{W}_i, b) = \mathbf{W}_i^T \mathbf{X}_i + b = \mathbf{W}_i \cdot \mathbf{X}_i + b$$

$$= \sum_{j=1}^{n} W_i^j X_i^j + b, \quad (12)$$

TABLE 11: Convolutional Neural Network.

Reference	Descriptor	Image type	Number of images	Key findings
Jiang et al. [89]	(1) Global Features	Mammogram	—	(1) Image preprocessing was performed to enhance tissue characteristics. (2) Transfer learning was performed and obtained AUC was 0.88 whereas when the system learned from scratch, the best ROC is 0.82.
Suzuki et al. [90]	(1) Global Features	Mammogram	198	(1) The achieved sensitivity 89.90%. (2) Transfer learning techniques have been utilized.
Qiu et al. [91]	(1) Global Features	Mammogram	270	(1) Average achieved Accuracy is 71.40%.
Samala et al. [92]	(1) Global Features	—	92	(1) They utilized Deep Learning CNN (DLCNN) and CNN models for classification. (2) The AUC of CNN and DLCNN model is 0.89 and 0.93, respectively.
Sharma and Preet [84]	(1) Global Features	Mammogram	607	(1) Transfer learning and ensemble techniques utilized. (2) When using ensemble techniques the soft voting method has been used. (3) The best ROC score is 0.86.
Kooi et al. [93]	(1) Global and Local features	Mammogram	44090	(1) Transfer learning method utilized (VGG model).
Geras et al. [94]	(1) Global Features	Mammogram	102800	(1) They investigated the relation of the Accuracy with the database size and image size.
Arevalo et al. [82]	(1) Global Features	Mammogram	736	(1) The best ROC value was 0.822.

TABLE 12: Logic Based.

Reference	Descriptor	Image type	Number of images	Key findings
Beura et al. [95]	(1) Two-dimensional discrete orthonormal S-transform has been used for the feature extraction	Mammogram	—	(1) Achieved Accuracy and AUC values on MIAS database are 98.3%, 0.9985. (2) Achieved Accuracy and AUC values on DDSM database are 98.8%, 0.9992.
Diz et al. [96]	(1) GLCM (2) GLRLM	Mammogram	410	(1) Their achieved Accuracy value is 76.60% (2) Mean false positive value is 81.00%.
Zhang et al. [97]	(1) 133 features (mass based and content based)	Mammogram	400	(1) Computer model has been created which is able to find a location that was not detected by trainee.
Ahmad and Yusoff [98]	(1) Nine features selected	Biopsy	700	(1) Achieved Sensitivity, Specificity, and Accuracy are 75.00%, 70.00%, and 72.00%, respectively.
Paul et al. [99]	(1) Harlick texture feature	Histopathological	50	(1) Their achieved Recall and Precision are 81.13% and 83.50%.
Chen et al. [100]	(1) Dual-tree complex wavelet transform (DT-CWT) has been used for the feature extraction.	Mammogram	—	(1) Achieved Received Operating Curve (ROC) 0.764.
Zhang et al. [101]	(1) Curvelet Transform (2) GLCM (3) CLBP	Histopathological	50	(1) Random Subspace Ensemble (RSE) utilized. (2) Their achieved classification Accuracy is 95.22% where the previous Accuracy on this same database was 93.40%.

Table 13: Logic Based.

| Reference	Descriptor	Image type	Number of images	Key findings
Angayarkanni and Kamal [102]	(1) GLCM	Mammogram	322	(1) The Achieved Sensitivity and Accuracy are 93.40% and 99.50%, respectively.
Wang et al. [103]	(1) Horizontal Weighted Sum (2) Vertical Weighted Sum (3) Diagonal Weighted Sum (4) Grid Weighted Sum.	Mammogram	322	(1) Surrounding Region Dependence Method (SRDM) utilized for region detection. (2) Achieved True Positive Rate 90.00% and False Positive Rate 88.80%.
Tambasco Bruno et al. [104]	(1) Curvelet Transform (2) LBP	Mammogram Histopathological	—	(1) ANOVA method utilized for feature prioritization. (2) When they use RF algorithm on Mammogram (DDSM) dataset, obtained Accuracy and ROC are 79.00% and 0.89.
Muramatsu et al. [105]	(1) Radial Local Ternary Pattern (RLTP)	Mammogram	376	(1) Textural features have been extracted from the regions of interest (ROIs) using RLTP. (2) They claimed that the RLTP feature provides better performance than the rotation invariant patterns.
Dong et al. [106]	(1) NRL margin gradient (2) Gray-level histogram (3) Pixel value fluctuation	Mammogram	—	(1) Chain code utilized for extraction of regions of interest (ROIs). (2) Rough-Set method utilized to enhance the ROIs. (3) Their achieved ROC value is 0.947 and obtained Matthews Correlation (MCC) is 0.8652.
Piantadosi et al. [107]	(1) Local Binary Pattern-Three Orthogonal Projections (LBP-TOP)	Mammogram	—	(1) Their achieved Accuracy, Sensitivity, and Specificity values are 84.60%, 80.00%, and 90.90%.
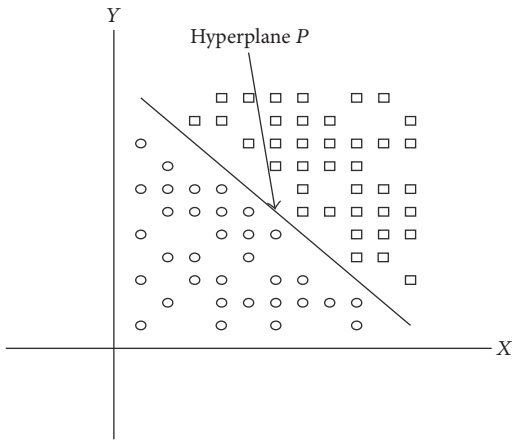
Figure 17: SVM finds the hyperplane which separates two classes.

where $\mathbf{W}_i, \mathbf{X}_i \in \mathscr{R}^n$. As the training data are linearly separable no training data will satisfy the condition

$$d\left(\mathbf{X}_i, \mathbf{W}_i, b\right) = 0. \tag{13}$$

To control the separability, we consider the following inequalities:

$$d\left(\mathbf{X}_i, \mathbf{W}_i, b\right) \geq 1 \quad \text{for } \omega_i = +1,$$
$$d\left(\mathbf{X}_i, \mathbf{W}_i, b\right) < 1 \quad \text{for } \omega_i = -1. \tag{14}$$

Sometime it is very difficult to find the perfect hyperplane which can separate the data, but if we transform the data into a higher dimension the data may be easily separable. To separate this kind of data, a kernel function can be introduced.

Kernel Methods. Assume a transformation ϕ such that it transforms the dataset $\mathbf{X}^1 \in \mathscr{R}^n$ into dataset $\mathbf{X}^2 \in \mathscr{R}^m$ where $m > n$. Now train the linear SVM on the dataset $\mathbf{X}^2$ to get a new classifier F_{SVM}.

A kernel ϕ effectively computes a dot product in a higher-dimensional space $\mathscr{R}^m$. For $\{\mathbf{x}_i, \mathbf{x}_j\} \in \mathscr{R}^N$, $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i, \mathbf{x}_j) \rangle_m$ is an inner product of $\mathscr{R}^m$, where $\phi(\mathbf{x})$ transforms $\mathbf{x}$ to $\mathscr{R}^m$. Consider $\{\mathbf{x}_i, \mathbf{x}_j\} \in \mathscr{R}^n$; then we can define the kernel as follows:

(i) Radial basis function kernel (rbf): $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma| < \phi(\mathbf{x}_i - \mathbf{x}_j) > |^2)$.

(ii) Polynomial kernel (polynomial): $K(\mathbf{x}_i, \mathbf{x}_j) = (\langle \phi(\mathbf{x}_i \cdot \mathbf{x}_j) \rangle + r)^d$.

(iii) Sigmoid kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\langle \phi(\mathbf{x}_i, \mathbf{x}_j) \rangle + r)$.

(iv) Linear kernel (linear): $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i, \mathbf{x}_j) \rangle$.

The advantage of the kernel method for breast cancer image classification using an SVM was first introduced by El-Naqa et al. [156]. They classify Microcalcification clusters in mammogram images (76 images were utilized for the

TABLE 14: SVM for breast image classification (Page-1).

Reference	Descriptor	Image type	Number of images	Key findings
Malik et al. [108]	(1) Speed of sound (2) Attenuation image vector (3) Reflection image vector	QTUS	—	(1) Glands, fat, skin, and connective tissue have been classified. (2) Both linear and nonlinear SVM classifier have been utilized. (3) Their experiment obtained 85.20% Accuracy.
Chang et al. [109]	(1) Textural features such as (i) Autocorrelation Coefficient (ii) Autocovariance Coefficient	Ultrasound	250	(1) Benign and malignant images have been classified. (2) Accuracy, Sensitivity, Specificity, positive predictive values, and negative predictive value are 85.60%, 95.45%, 77.86%, 77.21%, and 95.61%, respectively.
Akbay et al. [110]	(1) 52 features have been extracted	Mammogram	—	(1) Microcalcification (MC) Classification Accuracy 94.00%
Levman et al. [111]	(1) Relative Signal Intensities (2) Derivative of Signal Intensities (3) Relative Signal Intensities and their derivatives in one vector (4) (i) Maximum of signal intensity enhancement; (ii) time of maximum enhancement; (iii) time of maximum washout	MRI	76	(1) Benign and malignant lesions are investigated. (2) Linear kernel, a polynomial kernel, and a radial basis function kernel utilized along with the SVM method for the breast image classification.
de Oliveira Martins et al. [112]	(1) Ripley's K function	Mammogram	390	(1) Benign and malignant image classification. (2) The achieved Accuracy, Sensitivity, and Specificity are 94.94%, 92.86%, and 93.33%, respectively.

experiment where the total number of MCs was 1120). They utilized the SVM method along with the Gaussian kernel as well as the polynomial kernel. In 2003, Chang et al. classified a set of sonography images using SVM techniques where they consider that the image is surrounded by pickle noise [157], where the database contains 250 images. Their achieved Accuracy was 93.20%. A total of thirteen features, including shape, law, and gradient features, were utilized along with SVM and a Gaussian kernel for the mammogram image classification. They performed their operation on 193 mammogram images and achieved 83.70% sensitivity and 30.20% False Positive Rate [158]. SVM has been combined with the NN method by B. Sing et al. for ultrasound breast image classification where the database contained a total of 178 images. They performed a hybrid feature selection method to select the best features [159].

A breast ultrasound image is always very complex in nature. The Multiple Instance Learning (MIL) algorithm has been first used along with SVM for the breast image classification by [176], and their obtained Accuracy was 91.07%. The Concentric Circle BOW feature extraction method was utilized to extract the features and later the SVM method was used for breast image classification [177]. Their achieved Accuracy is 88.33% when the dimension of the features was

1000. A Bag of Features has been extracted from histopathological images (using SIFT and DCT) and using SVM for classification by Mhala and Bhandari [178]. The experiment is performed on a database which contains 361 images, where 119 images are normal, 102 images are ductal carcinoma in situ, and the rest of the images are invasive carcinoma. Their experiment achieved 100.00% classification Accuracy for ductal carcinoma in situ, 98.88% classification Accuracy for invasive carcinoma, and 100.00% classification Accuracy for normal image classification. A mammogram (DDSM) image database has been classified by Hiba et al. [179] by SVM along with the Bag of Feature method. Firstly the authors extract LBP and quantize the binary pattern information for feature extraction. Their obtained Accuracy was 91.25%.

Along with the above-mentioned work different breast image databases have been analyzed and classified using SVM. We have summarized some of the work related to SVM in Tables 14, 15, and 16.

3.1.5. Bayesian. A Bayesian classifier is a statistical method based on Bayes theorem. This method does not follow any explicit decision rule; however it depends on estimating probabilities. The Naive Bayes method can be considered one of the earlier Bayesian learning algorithms.

TABLE 15: SVM for breast image classification.

Reference	Descriptor	Image type	Number of images	Key findings
Zhang et al. [122]	(1) Fractional Fourier transform information utilized as features	Mammogram	200	(1) They selected ROI for avoiding redundant complexity. (2) When SVM and Principal Component Analysis were used together the achieved Accuracy, Sensitivity and Specificity are 92.16 ± 3.60%, 92.10 ± 2.75% and 92.22 ± 4.16% respectively.
Shirazi and Rashedi [123]	(1) GLCM	Ultrasound	322	(1) ROI extracted for reducing redundant complexity. (2) SVM and Mixed Gravitational Search Algorithm (MGSA) used together for feature reduction. (3) The achieved Accuracy 86.00%; however SVM with MGSA method achieved 93.10% Accuracy.
Sewak et al. [124]	(1) Radius, perimeter, area, compactness, smoothness, concavity, concave points, symmetry, fractal dimension, and texture of nuclei calculated	Biopsies	569	(1) Achieved Accuracy, Sensitivity, and Specificity are 99.29%, 100.00%, and 98.11%, respectively.
Dheeba and Tamil Selvi [125]	(1) The laws texture features utilized	Mammogram	322	(1) The achieved Accuracy is 86.10%.

TABLE 16: SVM for breast image classification.

Reference	Descriptor	Image type	Number of images	Key findings
Taheri et al. [126]	(1) Intensity information (2) Value of detected corner (3) Energy	Mammogram	600	(1) Classified images into normal and abnormal images. (2) Removing unwanted objects from the images for reducing the redundancy and computational complexity. (3) Achieved Precision and Recall rates are 96.80% and 92.5%, respectively.
Tan et al. [127]	(1) Shape, fat, presence of calcification texture, spiculation, Contrast, Isodensity type features selected (2) Total number of features 181	Mammogram	1200	(1) Features have been selected from the region of interest. (2) They utilized the radial basis function (RBF) for their analysis. (3) The Sequential Forward Floating Selection (SFFS) method utilized for the feature selection. (4) The area under the receiver operating characteristic curve was (AUC) = 0.805 ± 0.012.
Kavitha and Thyagharajan [128]	(1) Histogram of the intensity has been used as a statistical feature. (2) 2D Gabor filter utilized for the textural feature extraction (3) Clinical features extracted from the database directly	Mammogram	322	(1) When using SVM with the linear kernel the obtained Accuracy, Sensitivity, and Specificity are 98%, 100%, and 96%, respectively. (2) When using weighted feature SVM with weights the obtained Accuracy, Sensitivity, and Specificity are 90%, 100% and 75%, respectively.

The Naive Bayes (NB) method works on the basis of the Bayes formula, where each of the features is considered statistically independent. Consider a dataset with m samples, with each sample containing a feature vector $\mathbf{x}^k$ with n features [180] and belonging to a particular class c_k. According to the NB formula, the probability of the particular class c_k with the conditional vector $\mathbf{x}^k$ is represented as

$$P\left(c_k \mid \mathbf{x}^k\right) = \frac{P\left(\mathbf{x}^k \mid c_k\right) P\left(c_k\right)}{P\left(\mathbf{x}^k\right)}. \tag{15}$$

Applying the chain rule

$$P\left(\mathbf{x}_1^k, \mathbf{x}_2^k, \mathbf{x}_3^k, \ldots, \mathbf{x}_n^k \mid c_k\right) = \prod_{i=1}^{n} \mathbf{P}\left(\mathbf{x}_i^k \mid c_k\right). \tag{16}$$

The NB theorem considers all the features independently which can be represented as

$$\bar{c} = \arg\max_{k \in 1 \cdots m} P\left(c_k\right) \prod_{i=1}^{n} P\left(\mathbf{x}_i^k \mid c_k\right). \tag{17}$$

TABLE 17: Bayesian classifier.

Reference	Descriptor	Image type	Number of images	Key findings
Kendall and Flynn [129]	(1) Features extracted using DCT method.	Mammogram		(1) Bayesian classifier obtained 100.00% sensitivity with 64.00% specificity.
Oleksyuk et al. [130]		—	—	(1) Bayesian method obtained 86.00% with 80.00% specificity.
Burling-Claridge et al. [131]	(1) Statistical and LBP features extracted.	Mammogram	322/410	(1) Bayesian method obtained 67.07 ± 0.73% and 67.61 ± 0.83% Accuracy on MIAS and Inbreast image datasets (using statistical features). (2) Bayesian method obtained 62.86 ± 0.70% and 51.99 ± 1.28% Accuracy on MIAS and Inbreast image datasets (using LBP).
Raghavendra et al. [132]	(1) Gabor wavelet transform utilized for feature extraction.	Mammogram	690	(1) Locality Sensitive Discriminant Analysis (LSDA) for the data reduction. (2) NB obtained 84.34% Accuracy and 83.69% Sensitivity with 90.86% Specificity.
Pérez et al. [133]	(1) 23 features utilized.	Mammogram	—	(1) UFilter feature selection methods utilized and its efficiency verified by Wilcoxon statistical test.
Rashmi et al. [134]	(1) 10 features utilized.	—	—	(1) Benign and malignant tumors have been classified.
Gatuha and Jiang [135]	(1) 10 features utilized.	—	—	(1) They built an android based benign and malignant tumor classifier. (2) Their obtained Accuracy is 96.4%

The NB method is very easy to construct and very first to predict the data. This method can also utilize the kernel method. However, for a large dataset and continuous data, this method has very poor performance. NB can be classified into the following subclasses:

   (i) Gaussian Naive Bayes

   (ii) Multinomial Naive Bayes

   (iii) Bernoulli Naive Bayes.

One of the constraints of the NB classifier is that it considers that all the features are conditionally independent. A Bayesian Network is another Bayesian classifier which can overcome this constraint [181, 182]. The literature shows that the Bayesian classifier method is not utilized much for breast image classification. In 2003 Butler et al. used NB classifier for X-ray breast image classification [183]. They extracted features from the low-level pixels. For all feature combinations they obtained more than 90.00% Accuracy. Bayesian structural learning has been utilized for a breast lesion classifier by Fischer et al. [184]. Soria et al. [185] classify a breast cancer dataset utilizing C4.5, multilayered perceptron, and the NB algorithm using WEKA software [186]. They conclude that the NB method gives better performance than the other two methods in that particular case. They also compared their results with the Bayes classifier output. Some other research on the Bayes classifier and breast image classification has been summarized in Tables 17 and 18.

3.2. Performance Based on Unsupervised Learning. This learning algorithm does not require any prior knowledge about the target. The main goal of the unsupervised learning is to find the hidden structure and relations between the different data [187] and distribute the data into different clusters. Basically clustering is a statistical process where a set of data points is partitioned into a set of groups, known as a cluster. The K-means algorithm is a clustering algorithm proposed by [188]. Interestingly, unsupervised learning can be utilized as preprocessing step too.

   (i) In the K-means algorithm, firstly assign K centroid points. Suppose that we have n feature points x_i where $i \in \{1, \ldots, n\}$. The objective of the K-means algorithm is to find positions μ_i, where $i \in 1, \ldots, K$ that minimize the data points to the cluster by solving

$$\arg\min_{x \in c_i} \sum_{i=1}^{K} \sum_{x \in c_i} d(x, \mu_i) = \arg\min_{x \in c_i} \sum_{i=1}^{K} \sum_{x \in c_i} \|x - \mu_i\|^2. \quad (18)$$

   (ii) Self-Organizing Map (SOM): SOM is another popular unsupervised classifier, proposed by Kohonen et al. [189–191]. The main idea of the SOM method is to reduce the dimension of the data and represent those dimensionally reduced data by a map architecture, which provides more visual information.

   (iii) Fuzzy C-Means Clustering (FCM): the FCM algorithm cluster databased on the value of a membership function is proposed by [192] and improved by Bezdek [193].

The history of using unsupervised learning for breast image classification is a long one. In 2000, Cahoon et al. [194] classified mammogram breast images (DDSM database) in an unsupervised manner, utilizing the K-NN clustering and Fuzzy C-Means (FCM) methods. Chen et al. classified a set of breast images into benign and malignant classes [164].

TABLE 18: Bayesian classifier.

Reference	Descriptor	Image type	Number of images	Key findings
Benndorf et al. [136]	(1) BI-RADS features utilized.	—	2766	(1) For the training data the AUC value is 0.959 for the inclusive model, whereas AUC value is 0.910 for the descriptor model.
Rodríguez-López and Cruz-Barbosa [137]	(1) Eight image feature nodes utilized.	—	—	(1) NB model obtained 79.00% Accuracy, 80.00% Sensitivity.
Nugroho et al. [138]	(1) Eight image feature nodes utilized.	Mammogram	—	(1) Naive Bayes model along with SMO; obtained ROC value is 0.903. (2) Bayesian Network model along with SMO; obtained Accuracy was 83.68%.
Rodríguez-López and Cruz-Barbosa [139]	(1) Eight image features have been utilized.	—	231	(1) Bayesian Network model obtained 82.00% Accuracy, 80.00% Sensitivity, and 83.00% Specificity when they utilized only three features.
Shivakumari et al. [140]		—	231	(1) Analyze the Ljubljana breast image dataset. (2) NB algorithm along with feature ranking techniques; the best achieved Accuracy was 81.46%.
Rodríguez-López and Cruz-Barbosa [141]	(1) Seven different clinical features extracted.	Mammogram	690	(1) Obtained Accuracy, Sensitivity, and Specificity are 82.00%, 80.00%, and 83.00%, respectively.

TABLE 19: K-means Cluster Algorithm and Self-Organizing Map for breast image classification.

Reference	Descriptor	Image type	Number of images	Key findings
Moftah et al. [142]	(1) Intensity distribution used as feature.	MRI	—	(1) Three types of evaluation measures performed: (a) Accuracy, (b) feature based, (c) shape based measure. (2) This can classify the data as well as identify the target. (3) The obtained best Accuracy of the segmented ROI is 90.83%
Lee et al. [143]	(1) 1734 signal patterns.	MRI	322	(1) Available signal patterns have been classified into 10 classes.
Dalmiya et al. [144]	(1) Discrete Wavelet Transform.	Mammogram	—	(1) Cancer tumor masses have been segmented.
Elmoufidi et al. [145]	(1) Local Binary Pattern.	Mammogram	322	(1) Image enhancing. (2) Generation of number of clusters (3) Detection of regions of interest. (4) Mean detection of regions of interest is 85.00%.
Samundeeswari et al. [146]		Ultrasound	—	(1) Utilizing ant colony and regularization parameters. (2) This method obtained 96.00% similarity between segmented and reference tumors.
Rezaee [147]	Discrete Wavelet Transform.	Mammogram	120	(1) Early detection of tumors from the breast image. (2) Tumor detection Accuracy 92.32%, Sensitivity 90.24%.
Chandra et al. [148]	(1) Gray intensity values.	Mammogram	—	(1) Mammogram image has been clustered using SOM along with the Quadratic Neural Network.

They utilized a SOM procedure to perform this classification operation. They collected 24 autocorrelation textural features and used a 10-fold validation method. Markey et al. utilized the SOM method for BIRADS image classification of 4435 samples [195]. Tables 19 and 20 summarize the breast image classification performance based on K-means algorithm and SOM method.

3.3. *Performance Based on Semisupervisor.* The working principle of semisupervised learning lies in between supervised and unsupervised learning. For the semisupervised learning a few input data have an associated target and large amounts of data are not labeled [196]. It is always very difficult to collect the labeled data. Few data such as speech or information scratched from the web are difficult to label. To classify

TABLE 20: *K*-means Cluster Algorithm and Self-Organizing Map for breast image classification.

Reference	Descriptor	Image Type	No. of Images	Key Findings
Lashkari and Firouzmand [160]		Thermogram	23	(1) Both FCM method and Adaboost method utilized separately to classify images. (2) For the classification purposes selected 23 features and also select the best features using feature selection algorithm. When they used the FCM method, the obtained Mean Accuracy was 75.00% whereas the Adaboost method Accuracy was 88.00%.
Nattkemper et al. [161]		MRI	—	(1) *K*-means algorithm as well as SM method utilized.
Slazar-Licea et al. [162].		$\cdots$	—	(1) Fuzzy *c*-means algorithm used.
Marcomini et al. [163]	(1) 24 morphological features	Ultrasound	144	(1) Minimizing noise using Wiener filter, equalized and Median filter (2) Obtained Sensitivity 100% and Specificity 78.00%.
Chen et al. [164]	(1) 24 autocorrelation texture features	Ultrasound	243	(1) Obtained ROC area 0.9357 ± 0.0152. Accuracy 85.60%, Specificity 70.80%.
Iscan et al. [165]	(1) Two-dimensional discrete cosine transform (2) 2D continuous wavelet transform	Ultrasound	—	(1) Automated threshold scheme introduce to increase the robustness of the SOM algorithm.

this kind of data semisupervised learning is very efficient. However lately this method has been utilized for the brats image classification too. Semisupervised learning can be classified as

(i) Graph Based (GB)

(ii) Semisupervised Support Vector Machine

(iii) Human Semisupervised Learning.

To the best of our knowledge, Li and Yuen have utilized GB semisupervised learning for biomedical image classification [197]. The kernel trick is applied along with the semisupervised learning method for breast image classification by Li et al. [198]. They performed their experiments on the Wisconsin Prognostic Breast Cancer (WPBC) dataset for the breast image classification. Ngadi et al. utilized both the SKDA (Supervised Kernel-Based Deterministic Annealing) and NSVC methods for mammographic image classification [199]. They performed their experiments on 961 images, where 53.60% of the images were benign and the rest of the images are malignant. Among the other utilized features they utilized BI-RADS descriptors as features. When they utilized the NSVC method they also utilized RBF, polynomial, and linear kernel. They found that the best Accuracy of 99.27% was achieved when they utilized linear kernels. Few studies have performed the breast image classification by semisupervised learning, as summarized in Tables 21 and 22.

4. Conclusion

Breast cancer is a serious threat to women throughout the world and is responsible for increasing the female mortality

rate. The improvement of the current situation with breast cancer is a big concern and can be achieved by proper investigation, diagnosis, and appropriate patient and clinical management. Identification of breast cancer in the earlier stages and a regular check of the cancer can save many lives. The status of cancer changes with time, as the appearance, distribution, and structural geometry of the cells are changing on a particular time basis because of the chemical changes which are always going on inside the cell. The changing structure of cells can be detected by analysing biomedical images which can be obtained by mammogram, MRI, and so forth techniques. However these images are complex in nature and require expert knowledge to perfectly analyze malignancy. Due to the nontrivial nature of the images the physician sometimes makes a decision which might contradict others. However computer-aided-diagnosis techniques emphasising the machine learning can glean a significant amount of information from the images and provide a decision based on the gained information, such as cancer identification, by classifying the images.

The contribution of machine learning techniques to image classification is a long story. Using some advanced engineering techniques with some modifications, the existing machine learning based image classification techniques have been used for biomedical image classification, specially for breast image classification and segmentation. A few branches of the machine learning based image classifier are available such as Deep Neural Network, Logic Based, and SVM. Except for deep-learning, a machine learning-based classifier largely depends on handcrafted feature extraction techniques such as statistical and structural information that depend on various mathematical formulations and theorize where they gain

TABLE 21: Semisupervised algorithm for breast image classification.

Reference	Descriptor	Image type	Number of images	Key finding
Cordeiro et al. [166]	(1) Zernike moments have been used for the feature extraction.	—	685	(1) Semisupervised Fuzzy GrowCut algorithm utilized. (2) For the fatty-tissue classification this method achieved 91.28% Accuracy.
Cordeiro et al. [167]	—	Mammogram	322	(1) Semisupervised Fuzzy GrowCut as well as the Fuzzy GrowCut algorithm utilized for tumors, region segmentation.
Nawel et al. [168]	—	—	—	(1) Semisupervised Support Vector Machine (S3VM) utilized. (2) This experiment shows impressive results on the DDSM database.
Zemmal et al. [169]	—	DDSM	—	(1) Transductive semisupervised learning technique using (TSVM) utilized for classification along with different features.
Zemmal et al. [170]	—	—	200	(1) Semisupervised Support Vector Machine (S3VM) utilized with various kernels.
Zemmal et al. [171]	(1) GLCM (2) Hu moments (3) Central Moments	Mammogram	—	(1) Transductive Semisupervised learning technique used for image classification. (2) This experiment shows impressive results on DDSM database.
Peikari et al. [172]	(1) Mean, Mode, Standard Deviation, Media, Skewness, Kurtosis	Histopathological	322	(1) The Ordering Points to Identify the Clustering Structure (OPTICS) method utilized for image classification [173].

TABLE 22: Semisupervised algorithm for breast image classification.

Reference	Descriptor	Image type	Number of images	Key findings
Zhu et al. [174]	(1) Relative local intensity (2) Shape irregularity (3) Orientation consistency	Ultrasound	144	(1) One important microenvironment inside the tumor is vasculature, which has been classified in this paper.
Liu et al. [175]	—	Ultrasound	—	(1) Iterated Laplacian regularization based semisupervised algorithm for robust feature selection (Iter-LR-CRFS) utilized. (2) The archived Accuracy and Sensitivity are $89.0 \pm 3.6\%$ and $91.0 \pm 5.2\%$.

object-specific information. They are further utilized as an input for an image classifier such as SVM and Logic Based, for the image classification.

This investigation finds that most of the conventional classifiers depend on prerequisite local feature extraction. The nature of cancer is always changing, so the dependencies on a set of local features will not provide good results on a new dataset. However the state-of-the art Deep Neural Networks, specially CNN, have recently advanced biomedical image classification due to the Global Feature extraction capabilities. As the core of the CNN model is the kernel, which gives this model the luxury of working with the Global Features, these globally extracted features allow the CNN model to extract more hidden structure from the images. This allows some exceptional results for breast cancer image classification. As the CNN model is based on the Global

Features, this kind of classifier model should be easy to adapt to a new dataset.

This paper also finds that the malignancy information is concentrated in the particular area defined as ROI. Utilizing only the ROI portions, information gathered from the segmented part of the data can improve the performance substantially. The recent development of the Deep Neural Network can also be utilized for finding the ROI and segmenting the data, which can be further utilized for the image classification.

For breast cancer patient care, the machine learning techniques and tools have been a tremendous success so far, and this success has gained an extra impetus with the involvement of deep-learning techniques. However the main difficulty of handling the current deep-learning based machine learning classifier is its computational complexity, which is much

higher than for the traditional method. The current research is focused on the development of the light DNN model so that both the computational and timing complexities can be reduced. Another difficulty of using the DNN based cancer image classifier is that it requires a large amount of training data. However the reinforcement of learning techniques and data augmentation has been largely adapted with the current CNN model, which can provide reliable outcomes. Our research finds that the current trend of machine learning is largely towards deep-learning techniques. Among a few other implications, the appropriate tools for designing the overall deep-learning model was the initial obligation for utilizing deep-learning based machine learning techniques. However some reliable software has been introduced which can be utilized for breast image classification. Initially it was difficult to implement a DNN based architecture in simpler devices; however due to cloud-computer based Artificial Intelligence techniques this issue has been overcome and DNN has already been integrated with electronic devices such as mobile phones. In future combining the DNN network with the other learning techniques can provide more-positive predictions about breast cancer.

Due to the tremendous concern about breast cancer, many research contributions have been published so far. It is quite difficult to summarize all the research work related to breast cancer image classification based on machine learning techniques in a single research article. However this paper has attempted to provide a holistic approach to the breast cancer image classification procedure which summarizes the available breast dataset, generalized image classification techniques, feature extraction and reduction techniques, performance measuring criteria, and state-of-the-art findings.

In a nutshell, the involvement of machine learning for breast image classification allows doctors and physicians to take a second opinion, and it provides satisfaction to and raises the confidence level of the patient. There is also a scarcity of expert people who can provide the appropriate opinion about the disease. Sometimes the patient might need to spend a long time waiting due to the lack of expert people. In this particular scenario the machine learning based diagnostic system can help the patient to receive the timely feedback about the disease which can improve the patient-management scenario.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] J. E. Skandalakis, *Embryology and Anatomy of the Breast*, Springer Berlin Heidelberg, Berlin, Heidelberg, Germany, 2009.

[2] M. A. Shampo and R. A. Kyle, "Karl theodore dussik—pioneer in ultrasound," *Mayo Clinic proceedings*, vol. 70, no. 12, p. 1136, 1995.

[3] O. H. Karatas and E. Toy, "Three-dimensional imaging techniques: a literature review," *European Journal of Dentistry*, vol. 8, no. 1, pp. 132–140, 2014.

[4] M. Lakrimi, A. M. Thomas, G. Hutton et al., "The principles and evolution of magnetic resonance imaging," *Journal of Physics: Conference Series*, vol. 286, no. 1, Article ID 012016, 2011.

[5] http://www.aihw.gov.au/acim-books/.

[6] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using Convolutional Neural Networks," in *Proceedings of the 2016 International Joint Conference on Neural Networks, IJCNN 2016*, pp. 2560–2567, Canada, July 2016.

[7] R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786–804, 1979.

[8] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 8, no. 6, pp. 460–473, 1978.

[9] T. Lindeberg, "Feature detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, no. 2, pp. 79–116, 1998.

[10] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of the 4th Alvey Vision Conference*, pp. 147–151, 1988.

[11] S. M. Smith and J. M. Brady, "SUSAN: a new approach to low level image processing," *International Journal of Computer Vision*, vol. 23, no. 1, pp. 45–78, 1997.

[12] E. Rosten and T. Drummond, "Fusing points and lines for high performance tracking," in *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV '05)*, vol. 2, pp. 1508–1515, Beijing, China, October 2005.

[13] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," *Proceedings of the 9th European Conference on Computer Vision (ECCV '06)*, vol. Part I, Springer-Verlag, pp. 430–443, 2006.

[14] R. Lenz, "Rotation-invariant operators and scale-space filtering," *Pattern Recognition Letters*, vol. 6, no. 3, pp. 151–154, 1987.

[15] R. Lakemond, S. Sridharan, and C. Fookes, "Hessian-based affine adaptation of salient local image features," *Journal of Mathematical Imaging and Vision*, vol. 44, no. 2, pp. 150–167, 2012.

[16] T. Lindeberg, "Scale selection properties of generalized scale-space interest point detectors," *Journal of Mathematical Imaging and Vision*, vol. 46, no. 2, pp. 177–210, 2013.

[17] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[18] W. N. J. Hj Wan Yussof and M. S. Hitam, "Invariant Gabor-based interest points detector under geometric transformation," *Digital Signal Processing*, vol. 25, no. 1, pp. 190–197, 2014.

[19] J.-M. Morel and G. Yu, "Asift: A new framework for fully affine invariant image comparison," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 438–469, 2009.

[20] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. II-257–II-263, Madison, WI, USA, June 2003.

[21] B. Zhang, Y. Jiao, Z. Ma, Y. Li, and J. Zhu, "An efficient image matching method using Speed Up Robust Features," in *Proceedings of the 11th IEEE International Conference on Mechatronics and Automation, IEEE ICMA 2014*, pp. 553–558, China, August 2014.

[22] B. Karasfi, T. S. Hong, A. Jalalian, and D. Nakhaeinia, "Speedup Robust Features based unsupervised place recognition for assistive mobile robot," in *Proceedings of the 2011 International*

Conference on Pattern Analysis and Intelligent Robotics, ICPAIR 2011, pp. 97–102, Malaysia, June 2011.

[23] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[24] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[25] T. Ojala, M. Pietikäinen, and T. Mäenpää, "A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification," in *Proceedings of the Second International Conference on Advances in Pattern Recognition (ICAPR '01)*, pp. 397–406, Springer-Verlag, London, UK, 2001.

[26] T. Ahonen, J. Matas, C. He, and M. Pietikäinen, *Rotation Invariant Image Description with Local Binary Pattern Histogram Fourier Features*, pp. 61–70, Springer Berlin Heidelberg, Berlin, Heidelberg, Germany, 2009.

[27] G. Zhao and M. Pietikäinen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 915–928, 2007.

[28] M. Calonder, V. Lepetit, M. Özuysal, T. Trzcinski, C. Strecha, and P. Fua, "BRIEF: computing a local binary descriptor very fast," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1281–1298, 2012.

[29] D. Gong, S. Li, and Y. Xiang, "Face recognition using the Weber Local Descriptor," in *Proceedings of the 1st Asian Conference on Pattern Recognition, ACPR 2011*, pp. 589–592, China, November 2011.

[30] J. Chen, S. Shan, C. He et al., "WLD: A robust local image descriptor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1705–1720, 2010.

[31] S. H. Davarpanah, F. Khalid, L. Nurliyana Abdullah, and M. Golchin, "A texture descriptor: BackGround Local Binary Pattern (BGLBP)," *Multimedia Tools and Applications*, vol. 75, no. 11, pp. 6549–6568, 2016.

[32] M. Heikkilä, M. Pietikäinen, and C. Schmid, *Description of Interest Regions with Center-Symmetric Local Binary Patterns*, pp. 58–69, Springer Berlin Heidelberg, Berlin, Heidelberg, Germany, 2006.

[33] G. Xue, L. Song, J. Sun, and M. Wu, "Hybrid center-symmetric local pattern for dynamic background subtraction," in *Proceedings of the 2011 12th IEEE International Conference on Multimedia and Expo (ICME '11)*, pp. 1–6, July 2011.

[34] H. Wu, N. Liu, X. Luo, J. Su, and L. Chen, "Real-time background subtraction-based video surveillance of people by integrating local texture patterns," *Signal, Image and Video Processing*, vol. 8, no. 4, pp. 665–676, 2014.

[35] L. Liu, P. Fieguth, G. Zhao, M. Pietikäinen, and D. Hu, "Extended local binary patterns for face recognition," *Information Sciences*, vol. 358-359, pp. 56–72, 2016.

[36] T. Mäenpää and M. Pietikäinen, "Classification with color and texture: jointly or separately?" *Pattern Recognition*, vol. 37, no. 8, pp. 1629–1640, 2004.

[37] G. Xue, J. Sun, and L. Song, "Dynamic background subtraction based on spatial extended center-symmetric local binary pattern," in *Proceedings of the 2010 IEEE International Conference on Multimedia and Expo, ICME 2010*, pp. 1050–1054, Singapore, July 2010.

[38] S. Liao, G. Zhao, V. Kellokumpu, M. Pietikäinen, and S. Z. Li, "Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes," in *Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2010*, pp. 1301–1306, USA, June 2010.

[39] C. Silva, T. Bouwmans, and C. Frélicot, "An extended center-symmetric local binary pattern for background modeling and subtraction in videos," in *Proceedings of the 10th International Conference on Computer Vision Theory and Applications (VISAPP '15)*, vol. 1, pp. 395–402, 2015.

[40] Y. Chen, L. Ling, and Q. Huang, "Classification of breast tumors in ultrasound using biclustering mining and neural network," in *Proceedings of the 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, CISP-BMEI 2016*, pp. 1787–1791, China, October 2016.

[41] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: A review of classification and combining techniques," *Artificial Intelligence Review*, vol. 26, no. 3, pp. 159–190, 2006.

[42] K. T. Rajakeerthana, C. Velayutham, and K. Thangavel, *Mammogram Image Classification Using Rough Neural Network*, pp. 133–138, Springer India, New Delhi, Indina, 2014.

[43] V. Lessa and M. Marengoni, *Applying Artificial Neural Network for the Classification of Breast Cancer Using Infrared Thermographic Images*, pp. 429–438, Springer International Publishing, Cham, Germany, 2016.

[44] S. Wan, H.-C. Lee, X. Huang et al., "Integrated local binary pattern texture features for classification of breast tissue imaged by optical coherence microscopy," *Medical Image Analysis*, vol. 38, pp. 104–116, 2017.

[45] S. M. L. de Lima, A. G. da Silva-Filho, and W. P. dos Santos, "Detection and classification of masses in mammographic images in a multi-kernel approach," *Computer Methods and Programs in Biomedicine*, vol. 134, pp. 11–29, 2016.

[46] C. Abirami, R. Harikumar, and S. Chakravarthy, "Performance analysis and detection of micro calcification in digital mammograms using wavelet features," in *Proceedings of the International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET '16)*, pp. 2327–2331, Chennai, India, March 2016.

[47] N. El Atlas, A. Bybi, and H. Drissi, "Features fusion for characterizing INBREAST-database masses," in *Proceedings of the 2nd International Conference on Electrical and Information Technologies, ICEIT 2016*, pp. 374–379, Morocco, May 2016.

[48] H. Alharbi, G. Falzon, and P. Kwan, "A novel feature reduction framework for digital mammogram image classification," in *Proceedings of the 3rd IAPR Asian Conference on Pattern Recognition, ACPR 2015*, pp. 221–225, Malaysia, November 2016.

[49] W. Peng, R. V. Mayorga, and E. M. A. Hussein, "An automated confirmatory system for analysis of mammograms," *Computer Methods and Programs in Biomedicine*, vol. 125, pp. 134–144, 2016.

[50] A. Jalalian, S. Mashohor, R. Mahmud, B. Karasfi, M. Iqbal Saripan, and A. R. Ramli, "Computer-assisted diagnosis system for breast cancer in computed tomography laser mammography (ctlm)," *Journal of Digital Imaging*, pp. 1–16, 2017.

[51] H. Li, X. Meng, T. Wang, Y. Tang, and Y. Yin, "Breast masses in mammography classification with local contour features," *Biomedical Engineering Online*, vol. 16, no. 1, 44 pages, 2017.

[52] D.-R. Chen, R.-F. Chang, and Y.-L. Huang, "Computer-aided diagnosis applied to US of solid breast nodules by using neural networks," *Radiology*, vol. 213, no. 2, pp. 407–412, 1999.

[53] D.-R. Chen, R.-F. Chang, Y.-L. Huang, Y.-H. Chou, C.-M. Tiu, and P.-P. Tsai, "Texture analysis of breast tumors on sonograms," *Seminars in Ultrasound, CT and MRI*, vol. 21, no. 4, pp. 308–316, 2000.

[54] N. D. Marom, L. Rokach, and A. Shmilovici, "Using the confusion matrix for improving ensemble classifiers," in *Proceedings of the 2010 IEEE 26th Convention of Electrical and Electronics Engineers in Israel, IEEEI 2010*, pp. 555–559, Israel, November 2010.

[55] S. B. Kotsiantis, "Supervised machine learning: a review of classification techniques," in *Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*, pp. 3–24, 2007.

[56] F. Rosenblatt, *The Perceptron: A Perceiving and Recognizing Automaton*, Cornell Aeronautical Laboratory, Buffalo, New York, USA, 1957.

[57] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.

[58] R. Hecht-Nielsen, "Neural networks for perception," in *ch. Theory of the Backpropagation Neural Network*, vol. 2, pp. 65–93, Harcourt Brace Co, Orlando, FL, USA, 1992.

[59] J. Li, J. H. Cheng, J. Y. Shi, and F. Huang, "Brief introduction of back propagation (BP) neural network algorithm and its improvement," in *Advances in Computer Science and Information Engineering—Volume 2*, D. Jin and S. Lin, Eds., vol. 169 of *Advances in Intelligent and Soft Computing*, pp. 553–558, Springer, Berlin, Germany, 2012.

[60] A. Dawson, R. Austin Jr., and D. Weinberg, "Nuclear grading of breast carcinoma by image analysis: Classification by multivariate and neural network analysis," *American Journal of Clinical Pathology*, vol. 95, Supplement 1, no. 4, pp. S29–S37, 1991.

[61] D.-R. Chen, R.-F. Chang, W.-J. Kuo, M.-C. Chen, and Y.-L. Huang, "Diagnosis of breast tumors with sonographic texture analysis using wavelet transform and neural networks," *Ultrasound in Medicine & Biology*, vol. 28, no. 10, pp. 1301–1310, 2002.

[62] S. D. De S. Silva, M. G. F. Costa, W. C. De A. Pereira, and C. F. F. C. Filho, "Breast tumor classification in ultrasound images using neural networks with improved generalization methods," in *Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '15)*, pp. 6321–6325, Italy, August 2015.

[63] I. Saritas, "Prediction of breast cancer using artificial neural networks," *Journal of Medical Systems*, vol. 36, no. 5, pp. 2901–2907, 2012.

[64] E. López-Meléndez, L. D. Lara-Rodríguez, E. López-Olazagasti, B. Sánchez-Rinza, and E. Tepichin-Rodríguez, "BICAD: Breast image computer aided diagnosis for standard BIRADS 1 and 2 in calcifications," in *Proceedings of the 22nd Annual International Conference on Electronics, Communications and Computers, CONIELECOMP 2012*, pp. 190–195, Mexico, February 2012.

[65] https://github.com/BVLC/caffe.

[66] Y. Jia, E. Shelhamer, J. Donahue et al., "Caffe: convolutional architecture for fast feature embedding," *CoRR*, 2014.

[67] http://torch.ch/.

[68] http://www.vlfeat.org/matconvnet/.

[69] A. Vedaldi and K. Lenc, "Matconvnet - convolutional neural networks for MATLAB," *CoRR*, 2014.

[70] http://deeplearning.net/software/theano/.

[71] J. Bergstra, O. Breuleux, F. Bastien et al., "Theano: A cpu and gpu math compiler in python," in *Proceedings of the 9th Python in Science Conference*, pp. 3–10, 2010.

[72] https://www.tensorflow.org/.

[73] https://github.com/Microsoft/CNTK.

[74] https://keras.io.

[75] https://github.com/ml4j.

[76] http://ceit.aut.ac.ir/keyvanrad/DeeBNet.

[77] M. A. Keyvanrad and M. M. Homayounpour, "A brief survey on deep belief networks and introducing a new object oriented MATLAB toolbox (deebnet)," *CoRR*, vol. abs/1408.3264, 2014.

[78] C. Y. Wu, S.-C. B. Lo, M. T. Freedman, A. Hasegawa, R. A. Zuurbier, and S. K. Mun, "Classification of microcalcifications in radiographs of pathological specimen for the diagnosis of breast cancer," in *Proceedings of the Medical Imaging*, pp. 630–641, SPIE. Digital Library, Newport Beach, CA, USA, 1994.

[79] B. Sahiner, H.-P. Chan, N. Petrick et al., "Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images," *IEEE Transactions on Medical Imaging*, vol. 15, no. 5, pp. 598–610, 1996.

[80] S.-C. B. Lo, H. Li, Y. Wang, L. Kinnard, and M. T. Freedman, "A multiple circular path convolution neural network system for detection of mammographic masses," *IEEE Transactions on Medical Imaging*, vol. 21, no. 2, pp. 150–158, 2002.

[81] P. Fonseca, J. Mendoza, J. Wainer et al., "Automatic breast density classification using a convolutional neural network architecture search procedure," in *Proceedings of the SPIE Medical Imaging Symposium 2015: Computer-Aided Diagnosis*, vol. 9414, pp. 941428–941428–8, USA, February 2015.

[82] J. Arevalo, F. A. González, R. Ramos-Pollán, J. L. Oliveira, and M. A. Guevara Lopez, "Representation learning for mammography mass lesion classification with convolutional neural networks," *Computer Methods and Programs in Biomedicine*, vol. 127, pp. 248–257, 2016.

[83] H. Su, F. Liu, Y. Xie, F. Xing, S. Meyyappan, and L. Yang, "Region segmentation in histopathological breast cancer images using deep convolutional neural network," in *Proceedings of the 12th IEEE International Symposium on Biomedical Imaging, ISBI 2015*, pp. 55–58, USA, April 2015.

[84] K. Sharma and B. Preet, "Classification of mammogram images by using CNN classifier," in *Proceedings of the 5th International Conference on Advances in Computing, Communications and Informatics, ICACCI 2016*, pp. 2743–2749, India, September 2016.

[85] H. Rezaeilouyeh, A. Mollahosseini, and M. H. Mahoor, "Microscopic medical image classification framework via deep learning and shearlet transform," *Journal of Medical Imaging*, vol. 3, no. 4, Article ID 044501, 2016.

[86] A. Albayrak and G. Bilgin, *Mitosis Detection Using Convolutional Neural Network Based Features*, pp. 335–340, 2017.

[87] Z. Jiao, X. Gao, Y. Wang, and J. Li, "A deep feature based framework for breast masses classification," *Neurocomputing*, vol. 197, pp. 221–231, 2016.

[88] M. Zejmo, M. Kowal, J. Korbicz, and R. Monczak, "Classification of breast cancer cytological specimen using convolutional neural network," *Journal of Physics: Conference Series*, vol. 783, no. 1, Article ID 012060, 2017.

[89] F. Jiang, H. Liu, S. Yu, and Y. Xie, "Breast mass lesion classification in mammograms by transfer learning," in *Proceedings of the 5th International Conference on Bioinformatics and Computational Biology (ICBCB '17)*, pp. 59–62, ACM, New York, NY, USA, 2017.

[90] S. Suzuki, X. Zhang, N. Homma et al., "Mass detection using deep convolutional neural network for mammographic computer-aided diagnosis," in *Proceedings of the 55th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE '16)*, pp. 1382–1386, Japan, September 2016.

[91] Y. Qiu, Y. Wang, S. Yan et al., "An initial investigation on developing a new method to predict short-term breast cancer risk based on deep learning technology," in *Proceedings of the Medical Imaging 2016: Computer-Aided Diagnosis*, SPIE. Digital Library, San Diego, California, USA, March 2016.

[92] R. K. Samala, H.-P. Chan, L. M. Hadjiiski, K. Cha, and M. A. Helvie, "Deep-learning convolution neural network for computer-aided detection of microcalcifications in digital breast tomosynthesis," in *Proceedings of the Medical Imaging 2016: Computer-Aided Diagnosis*, USA, March 2016.

[93] T. Kooi, G. Litjens, B. van Ginneken et al., "Large scale deep learning for computer aided detection of mammographic lesions," *Medical Image Analysis*, vol. 35, pp. 303–312, 2017.

[94] K. J. Geras, S. Wolfson, S. G. Kim, L. Moy, and K. Cho, "High-resolution breast cancer screening with multi-view deep convolutional neural networks," *CoRR*, vol. abs/1703.07047, 2017.

[95] S. Beura, B. Majhi, R. Dash, and S. Roy, "Classification of mammogram using two-dimensional discrete orthonormal S-transform for breast cancer detection," *Healthcare Technology Letters*, vol. 2, no. 2, pp. 46–51, 2015.

[96] J. Diz, G. Marreiros, and A. Freitas, *Using Data Mining Techniques to Support Breast Cancer Diagnosis*, Advances in Intelligent Systems and Computing, pp. 689–700, Springer International Publishing, Cham, Switzerland, 2015.

[97] J. Zhang, J. I. Silber, and M. A. Mazurowski, "Modeling false positive error making patterns in radiology trainees for improved mammography education," *Journal of Biomedical Informatics*, vol. 54, pp. 50–57, 2015.

[98] F. K. Ahmad and N. Yusoff, "Classifying breast cancer types based on fine needle aspiration biopsy data using random forest classifier," in *Proceedings of the 2013 13th International Conference on Intellient Systems Design and Applications (ISDA '13)*, pp. 121–125, Malaysia, December 2013.

[99] A. Paul, A. Dey, D. P. Mukherjee, J. Sivaswamy, and V. Tourani, *Regenerative Random Forest with Automatic Feature Selection to Detect Mitosis in Histopathological Breast Cancer Images*, vol. 9350 of *Lecture Notes in Computer Science*, pp. 94–102, Springer International Publishing, Cham, Switzerland, 2015.

[100] Z. Chen, M. Berks, S. Astley, and C. Taylor, *Classification of Linear Structures in Mammograms Using Random Forests*, Lecture Notes in Computer Science, pp. 153–160, Springer Berlin Heidelberg, Heidelberg, Germany, 2010.

[101] Y. Zhang, B. Zhang, and W. Lu, "Breast cancer classification from histological images with multiple features and random subspace classifier ensemble," in *Proceedings of the 2011 International Symposium on Computational Models for Life Sciences (CMLS '11)*, vol. 1371 of *AIP Conference Proceedings*, pp. 19–28, 2011.

[102] S. P. Angayarkanni and N. B. Kamal, "MRI mammogram image classification using ID3 algorithm," in *Proceedings of the IET Conference on Image Processing (IPR '12)*, pp. 1–5, IET, London, UK, July 2012.

[103] K. Wang, M. Dong, Z. Yang, Y. Guo, and Y. Ma, "Regions of micro-calcifications clusters detection based on new features from imbalance data in mammograms," in *Proceedings of the 2016 8th International Conference on Graphic and Image Processing (ICGIP '16)*, vol. 10225, pp. 102252C–102252C–6, SPIE. Digital Library, Tokyo, Japan, 2017.

[104] D. O. Tambasco Bruno, M. Z. Do Nascimento, R. P. Ramos, V. R. Batista, L. A. Neves, and A. S. Martins, "LBP operators on curvelet coefficients as an algorithm to describe texture in breast cancer tissues," *Expert Systems with Applications*, vol. 55, pp. 329–340, 2016.

[105] C. Muramatsu, T. Hara, T. Endo, and H. Fujita, "Breast mass classification on mammograms using radial local ternary patterns," *Computers in Biology and Medicine*, vol. 72, pp. 43–53, 2016.

[106] M. Dong, X. Lu, Y. Ma, Y. Guo, Y. Ma, and K. Wang, "An efficient approach for automated mass segmentation and classification in mammograms," *Journal of Digital Imaging*, vol. 28, no. 5, pp. 613–625, 2015.

[107] G. Piantadosi, R. Fusco, A. Petrillo, M. Sansone, and C. Sansone, *LBP-TOP for Volume Lesion Classification in Breast DCE-MRI*, pp. 647–657, Springer International Publishing, Cham, Switzerland, 2015.

[108] B. Malik, J. Klock, J. Wiskin, and M. Lenox, "Objective breast tissue image classification using Quantitative Transmission ultrasound tomography," *Scientific Reports*, vol. 6, no. 3, Article ID 38857, 2016.

[109] R.-F. Chang, W.-J. Wu, W. K. Moon, Y.-H. Chou, and D.-R. Chen, "Support vector machines for diagnosis of breast tumors on US images," *Academic Radiology*, vol. 10, no. 2, pp. 189–197, 2003.

[110] C. Akbay, N. G. Gençer, and G. Gençer, "CAD for detection of microcalcification and classification in Mammograms," in *Proceedings of the 2014 18th National Biomedical Engineering Meeting (BIYOMUT '14)*, pp. 1–4, Turkey, October 2014.

[111] J. Levman, T. Leung, P. Causer, D. Plewes, and A. L. Martel, "Classification of dynamic contrast-enhanced magnetic resonance breast lesions by support vector machines," *IEEE Transactions on Medical Imaging*, vol. 27, no. 5, pp. 688–696, 2008.

[112] L. de Oliveira Martins, E. C. da Silva, A. C. Silva, A. C. de Paiva, and M. Gattass, "Classification of Breast Masses in Mammogram Images Using Ripley's K Function and Support Vector Machine," in *Machine Learning and Data Mining in Pattern Recognition*, vol. 4571 of *Lecture Notes in Computer Science*, pp. 784–794, Springer Berlin Heidelberg, Berlin, Heidelberg, Germany, 2007.

[113] K. Fukushima, "Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological Cybernetics*, vol. 36, no. 4, pp. 193–202, 1980.

[114] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., pp. 1097–1105, Curran Associates, Inc., 2012.

[115] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," *CoRR*, vol. abs/1409.4842, 2014.

[116] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015.

[117] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *CoRR*, vol. abs/1512.00567, 2015.

[118] C. Szegedy, S. Ioffe, and V. Vanhoucke, "Inception-v4, inception-resnet and the impact of residual connections on learning," *CoRR*, vol. abs/1602.07261, 2016.

[119] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu et al., "Convolutional neural networks for medical image analysis: full training or fine tuning?" *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.

[120] Y. Liu, S. Zhou, and Q. Chen, "Discriminative deep belief networks for visual data classification," *Pattern Recognition*, vol. 44, no. 10-11, pp. 2287–2296, 2011.

[121] A. M. Abdel-Zaher and A. M. Eldeib, "Breast cancer classification using deep belief networks," *Expert Systems with Applications*, vol. 46, pp. 139–144, 2016.

[122] Y.-D. Zhang, S.-H. Wang, G. Liu, and J. Yang, "Computer-aided diagnosis of abnormal breasts in mammogram images by weighted-type fractional Fourier transform," *Advances in Mechanical Engineering*, vol. 8, no. 2, pp. 1–11, 2016.

[123] F. Shirazi and E. Rashedi, "Detection of cancer tumors in mammography images using support vector machine and mixed gravitational search algorithm," in *Proceedings of the 1st Conference on Swarm Intelligence and Evolutionary Computation (CSIEC '16)*, pp. 98–101, Iran, March 2016.

[124] M. Sewak, P. Vaidya, C.-C. Chan, and Z.-H. Duan, "SVM approach to breast cancer classification," in *Proceedings of the 2nd International Multi-Symposiums on Computer and Computational Sciences 2007 (IMSCCS '07)*, pp. 32–37, IEEE, Iowa City, IA, USA, August 2007.

[125] J. Dheeba and S. Tamil Selvi, "Classification of malignant and benign microcalcification using SVM classifier," in *Proceedings of the 2011 International Conference on Emerging Trends in Electrical and Computer Technology (ICETECT '11)*, pp. 686–690, India, March 2011.

[126] M. Taheri, G. Hamer, S. H. Son, and S. Y. Shin, "Enhanced breast cancer classification with automatic thresholding using SVM and Harris corner detection," in *Proceedings of the International Conference on Research in Adaptive and Convergent Systems (RACS '16)*, pp. 56–60, ACM, Odense, Denmark, October 2016.

[127] M. Tan, J. Pu, and B. Zheng, "Optimization of breast mass classification using sequential forward floating selection (SFFS) and a support vector machine (SVM) model," *International Journal for Computer Assisted Radiology and Surgery*, vol. 9, no. 6, pp. 1005–1020, 2014.

[128] S. Kavitha and K. K. Thyagharajan, "Features based mammogram image classification using weighted feature support vector machine," *Communications in Computer and Information Science*, vol. 270, no. II, pp. 320–329, 2012.

[129] E. J. Kendall and M. T. Flynn, "Automated breast image classification using features from its discrete cosine transform," *PLoS ONE*, vol. 9, no. 3, Article ID e91015, pp. 1–8, 2014.

[130] V. Oleksyuk, F. Saleheen, D. F. Caroline, S. A. Pascarella, and C.-H. Won, "Classification of breast masses using Tactile Imaging System and machine learning algorithms," in *Proceedings of the 2016 IEEE Signal Processing in Medicine and Biology Symposium (SPMB '16)*, pp. 1–4, USA, Dec 2016.

[131] F. Burling-Claridge, M. Iqbal, and M. Zhang, "Evolutionary algorithms for classification of mammographie densities using local binary patterns and statistical features," in *Proceedings of the 2016 IEEE Congress on Evolutionary Computation (CEC '16)*, pp. 3847–3854, Canada, July 2016.

[132] U. Raghavendra, U. Rajendra Acharya, H. Fujita, A. Gudigar, J. H. Tan, and S. Chokkadi, "Application of Gabor wavelet and Locality Sensitive Discriminant Analysis for automated identification of breast cancer using digitized mammogram images," *Applied Soft Computing*, vol. 46, pp. 151–161, 2016.

[133] N. P. Pérez, M. A. Guevara López, A. Silva, and I. Ramos, "Improving the Mann-Whitney statistical test for feature selection: an approach in breast cancer diagnosis on mammography," *Artificial Intelligence in Medicine*, vol. 63, no. 1, pp. 19–31, 2015.

[134] G. D. Rashmi, A. Lekha, and N. Bawane, "Analysis of efficiency of classification and prediction algorithms (Naïve Bayes) for Breast Cancer dataset," in *Proceedings of the 2015 International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT '15)*, pp. 108–113, IEEE, Mandya, India, December 2015.

[135] G. Gatuha and T. Jiang, "Android based Naive Bayes probabilistic detection model for breast cancer and Mobile Cloud Computing: Design and Implementation," *International Journal of Engineering Research in Africa*, vol. 21, pp. 197–208, 2016.

[136] M. Benndorf, E. Kotter, M. Langer, C. Herda, Y. Wu, and E. S. Burnside, "Development of an online, publicly accessible naive Bayesian decision support tool for mammographic mass lesions based on the American College of Radiology (ACR) BI-RADS lexicon," *European Radiology*, vol. 25, no. 6, pp. 1768–1775, 2015.

[137] V. Rodríguez-López and R. Cruz-Barbosa, "Improving bayesian networks breast mass diagnosis by using clinical data," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9116, pp. 292–301, 2015.

[138] K. A. Nugroho, N. A. Setiawan, and T. B. Adji, "Cascade generalization for breast cancer detection," in *Proceedings of the 2013 5th International Conference on Information Technology and Electrical Engineering (ICITEE '13)*, pp. 57–61, IEEE, Yogyakarta, Indonesia, October 2013.

[139] V. Rodríguez-López and R. Cruz-Barbosa, "On the breast mass diagnosis using Bayesian networks," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8857, pp. 474–485, 2014.

[140] S. Sivakumari, R. Praveena Priyadarsini, and P. Amudha, "Accuracy evaluation of C4.5 and Naïve Bayes classifiers using attribute ranking method," *International Journal of Computational Intelligence Systems*, vol. 2, no. 1, pp. 60–68, 2009.

[141] V. Rodríguez-López and R. Cruz-Barbosa, *Improving Bayesian Networks Breast Mass Diagnosis by Using Clinical Data*, pp. 292–301, Springer International Publishing, Cham, Switzerland, 2015.

[142] H. M. Moftah, A. T. Azar, E. T. Al-Shammari, N. I. Ghali, A. E. Hassanien, and M. Shoman, "Adaptive k-means clustering algorithm for MR breast image segmentation," *Neural Computing and Applications*, vol. 24, no. 7-8, pp. 1917–1928, 2014.

[143] S. H. Lee, J. H. Kim, K. G. Kim, S. J. Park, and W. K. Moon, *K-Means Clustering and Classification of Kinetic Curves on Malignancy in Dynamic Breast MRI*, pp. 2536–2539, Springer Berlin Heidelberg, Berlin, Heidelberg, Germany, 2007.

[144] S. Dalmiya, A. Dasgupta, and S. Kanti Datta, "Application of Wavelet based K-means Algorithm in Mammogram Segmentation," *International Journal of Computer Applications*, vol. 52, no. 15, pp. 15–19, 2012.

[145] A. Elmoufidi, K. El Fahssi, S. J. Andaloussi, and A. Sekkaki, "Detection of regions of interest in mammograms by using local binary pattern and dynamic K-means algorithm," *Orb Academic Publisher*, 2014.

[146] E. S. Samundeeswari, P. K. Saranya, and R. Manavalan, "Segmentation of Breast Ultrasound image using Regularized K-Means (ReKM) clustering," in *Proceedings of the 2016 IEEE International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET '16)*, pp. 1379–1383, India, March 2016.

[147] J. H. K. Rezaee, "Designing an algorithm for cancerous tissue segmentation using adaptive k-means cluttering and discrete wavelet transform," *Journal of Biomedical Physics and Engineering*, pp. 93–104, 2013.

[148] B. Chandra, S. Nath, and A. Malhothra, "Classification and clustering of breast cancer images," in *Proceedings of the International Joint Conference on Neural Networks 2006 (IJCNN '06)*, pp. 3843–3847, 2006.

[149] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, no. 1, pp. 81–106, 1986.

[150] J. R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.

[151] A. I. Pritom, M. A. R. Munshi, S. A. Sabab, and S. Shihab, "Predicting breast cancer recurrence using effective classification and feature selection technique," in *Proceedings of the 19th International Conference on Computer and Information Technology (ICCIT '16)*, pp. 310–314, December 2016.

[152] H. Asri, H. Mousannif, H. Al Moatassime, and T. Noel, "Using machine learning algorithms for breast cancer risk prediction and diagnosis," *Procedia Computer Science*, vol. 83, pp. 1064–1069, 2016.

[153] L. Breiman, "Arcing classifiers," *The Annals of Statistics*, vol. 26, no. 3, pp. 801–849, 1998.

[154] J. H. Friedman, "Stochastic gradient boosting," *Computational Statistics and Data Analysis*, vol. 38, Nonlinear Methods and Data Mining, no. 4, pp. 367–378, 2002.

[155] T. Chen and C. Guestrin, "Xgboost: a scalable tree boosting system," *CoRR*, vol. abs/1603.02754, 2016.

[156] I. El-Naqa, Y. Yang, M. N. Wernick, N. P. Galatsanos, and R. M. Nishikawa, "A support vector machine approach for detection of microcalcifications," *IEEE Transactions on Medical Imaging*, vol. 21, no. 12, pp. 1552–1563, 2002.

[157] R.-F. Chang, W.-J. Wu, W. K. Moon, and D.-R. Chen, "Improvement in breast tumor discrimination by support vector machines and speckle-emphasis texture analysis," *Ultrasound in Medicine & Biology*, vol. 29, no. 5, pp. 679–686, 2003.

[158] Y. Chu, L. Li, D. Goldgof, Y. Qiu, and R. A. Clark, "Classification of masses on mammograms using support vector machine," in *Proceedings of the Medical Imaging 2003: Image Processing*, pp. 940–948, USA, February 2003.

[159] B. K. Singh, K. Verma, A. Thoke, and J. S. Suri, "Risk stratification of 2D ultrasound-based breast lesions using hybrid feature selection in machine learning paradigm," *Measurement*, vol. 105, pp. 146–157, 2017.

[160] A. Lashkari and M. Firouzmand, "Early breast cancer detection in thermogram images using AdaBoost classifier and fuzzy C-Means clustering algorithm," *Middle East Journal of Cancer*, vol. 7, no. 3, pp. 113–124, 2016.

[161] T. W. Nattkemper, B. Arnrich, O. Lichte et al., "Evaluation of radiological features for breast tumour classification in clinical screening with machine learning methods," *Artificial Intelligence in Medicine*, vol. 34, no. 2, pp. 129–139, 2005.

[162] L. A. Salazar-Licea, J. C. Pedraza-Ortega, A. Pastrana-Palma, and M. A. Aceves-Fernandez, "Location of mammograms ROI's and reduction of false-positive," *Computer Methods and Programs in Biomedicine*, vol. 143, pp. 97–111, 2017.

[163] K. D. Marcomini, A. A. O. Carneiro, and H. Schiabel, "Application of artificial neural network models in segmentation and classification of nodules in breast ultrasound digital images," *International Journal of Biomedical Imaging*, vol. 2016, Article ID 7987212, 13 pages, 2016.

[164] D.-R. Chen, R.-F. Chang, and Y.-L. Huang, "Breast cancer diagnosis using self-organizing map for sonography," *Ultrasound in Medicine & Biology*, vol. 26, no. 3, pp. 405–411, 2000.

[165] Z. Iscan, Z. Dokur, and T. Ölmez, *Improved Incremental Self-Organizing Map for The Segmentation of Ultrasound Images*, pp. 293–302, Springer Netherlands, Dordrecht, Netherlands, 2007.

[166] F. R. Cordeiro, W. P. Santos, and A. G. Silva-Filho, "A semi-supervised fuzzy GrowCut algorithm to segment and classify regions of interest of mammographic images," *Expert Systems with Applications*, vol. 65, pp. 116–126, 2016.

[167] F. R. Cordeiro, W. P. Santos, and A. G. Silva-Filho, "Analysis of supervised and semi-supervised GrowCut applied to segmentation of masses in mammography images," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*, vol. 5, no. 4, pp. 297–315, 2017.

[168] Z. Nawel, A. Nabiha, D. Nilanjan, and S. Mokhtar, "Adaptive semi supervised support vector machine semi supervised learning with features cooperation for breast cancer classification," *Journal of Medical Imaging and Health Informatics*, vol. 6, no. 1, pp. 53–62, 2016.

[169] N. Zemmal, N. Azizi, and M. Sellami, "CAD system for classification of mammographic abnormalities using transductive semi supervised learning algorithm and heterogeneous features," in *Proceedings of the 12th International Symposium on Programming and Systems (ISPS '15)*, pp. 245–253, IEEE, Algiers, Algeria, April 2015.

[170] N. Zemmal, N. Azizi, N. Dey, and M. Sellami, "Adaptative S3VM semi supervised learning with features cooperation for breast cancer classification," *Journal of Medical Imaging and Health Informatics*, vol. 6, no. 4, pp. 957–967, 2016.

[171] N. Zemmal, N. Azizi, M. Sellami, and N. Dey, "Automated classification of mammographic abnormalities using transductive semi supervised learning algorithm," in *Proceedings of the Mediterranean Conference on Information & Communication Technologies 2015*, A. El Oualkadi, F. Choubani, and A. El Moussati, Eds., pp. 657–662, Springer International Publishing, Cham, 2016.

[172] M. Peikari, J. Zubovits, G. Clarke, and A. L. Martel, "Clustering analysis for semi-supervised learning improves classification performance of digital pathology," in *Proceedings of the International Workshop on Machine Learning in Medical Imaging. MICCAI 2015*, vol. 9352 of *Lecture Notes in Computer Science*, pp. 263–270, Springer International Publishing, Cham, Switzerland, 2015.

[173] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, "Optics: Ordering points to identify the clustering structure," *SIGMOD Rec*, vol. 28, pp. 49–60, June 1999.

[174] Y. Zhu, F. Li, T. J. Vadakkan et al., "Three-dimensional vasculature reconstruction of tumour microenvironment via local clustering and classification," *Interface Focus*, vol. 3, no. 4, 2013.

[175] X. Liu, J. Shi, S. Zhou, and M. Lu, "An iterated Laplacian based semi-supervised dimensionality reduction for classification of breast cancer on ultrasound images," in *Proceedings of the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '14)*, pp. 4679–4682, USA, August 2014.

[176] J. Ding, H. D. Cheng, J. Huang, J. Liu, and Y. Zhang, "Breast ultrasound image classification based on multiple-instance learning," *Journal of Digital Imaging*, vol. 25, no. 5, pp. 620–627, 2012.

[177] M. Pang, Y. Wang, and J. Li, "Dirichlet-based concentric circle feature transform for breast mass classification," in *Proceedings of the 27th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2015)*, vol. 2016, pp. 272–277, IEEE, Vietri sul Mare, Italy, November 2015.

[178] N. C. Mhala and S. H. Bhandari, "Improved approach towards classification of histopathology images using bag-of-features," in *Proceedings of the 2016 International Conference on Signal and Information Processing (IConSIP '16)*, IEEE, Vishnupuri, India, October 2016.

[179] C. Hiba, Z. Hamid, and A. Omar, "An improved breast tissue density classification framework using bag of features model," in *Proceedings of the 4th IEEE International Colloquium on Information Science and Technology, CiSt 2016*, pp. 405–409, Morocco, October 2016.

[180] P. Langley, W. Iba, and K. Thompson, "Analysis of Bayesian classifiers," in *Proceedings of the 10th National Conference on Artificial Intelligence (AAAI '92)*, pp. 223–228, AAAI Press, San Jose, Calif, USA, July 1992.

[181] A. Tosun, A. B. Bener, and S. Akbarinasaji, "A systematic literature review on the applications of Bayesian networks to predict software quality," *Software Quality Journal*, vol. 25, no. 1, pp. 273–305, 2017.

[182] J. Grover, "A Literature Review of Bayes' Theorem and Bayesian Belief Networks (BBN)," in *Strategic Economic Decision-Making*, vol. 9 of *SpringerBriefs in Statistics*, pp. 11–27, Springer New York, New York, NY, 2013.

[183] S. M. Butler, G. I. Webb, and R. A. Lewis, "A case study in feature invention for breast cancer diagnosis using X-ray scatter images," in *AI 2003: advances in artificial intelligence*, vol. 2903 of *Lecture Notes in Computer Science*, pp. 677–685, Springer, Berlin, Germany, 2003.

[184] E. A. Fischer, J. Y. Lo, and M. K. Markey, "Bayesian networks of BI-RADS/spl trade/ descriptors for breast lesion classification," in *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEMBS '04)*, vol. 26 IV, pp. 3031–3034, IEEE, San Francisco, CA, USA, Sept. 2004.

[185] D. Soria, J. M. Garibaldi, E. Biganzoli, and I. O. Ellis, "A comparison of three different methods for classification of breast cancer data," in *Proceedings of the 7th International Conference on Machine Learning and Applications (ICMLA '08)*, pp. 619–624, USA, December 2008.

[186] http://www.cs.waikato.ac.nz/ml/weka/.

[187] T. Masquelier and S. J. Thorpe, "Unsupervised learning of visual features through spike timing dependent plasticity," *PLoS Computational Biology*, vol. 3, no. 2, pp. 1–11, 2007.

[188] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 281–297, University of California Press, Berkeley, Calif, USA, 1967.

[189] T. Kohonen, M. R. Schroeder, and T. S. Huang, Eds., *Self-Organizing Maps*, Springer-Verlag New York, Secaucus, NJ, USA, 3rd edition, 2001.

[190] T. Kohonen, "Essentials of the self-organizing map," *Neural Networks*, vol. 37, Twenty-fifth Anniversay Commemorative Issue, pp. 52–65, 2013.

[191] T. Kohonen, "The Self-Organizing Map," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990.

[192] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *Journal of Cybernetics*, vol. 3, no. 3, pp. 32–57, 1973.

[193] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Kluwer Academic Publishers, Norwell, MA, USA, 1981.

[194] T. C. Cahoon, M. A. Sutton, and J. C. Bezdek, "Breast cancer detection using image processing techniques," in *Proceedings of the FUZZ-IEEE 2000: 9th IEEE International Conference on Fuzzy Systems*, pp. 973–976, May 2000.

[195] M. K. Markey, J. Y. Lo, G. D. Tourassi, and C. E. Floyd Jr., "Self-organizing map for cluster analysis of a breast cancer database," *Artificial Intelligence in Medicine*, vol. 27, no. 2, pp. 113–127, 2003.

[196] X. Zhu, "Semi-supervised learning literature survey," Tech. Rep., University of Wisconsin-Madison, 2005, Tech. Rep. 1530, Computer Sciences.

[197] C. H. Li and P. C. Yuen, "Semi-supervised Learning in Medical Image Database," in *Advances in Knowledge Discovery and Data Mining*, vol. 2035 of *Lecture Notes in Computer Science*, pp. 154–160, Springer Berlin Heidelberg, Berlin, Heidelberg, Germany, 2001.

[198] J.-B. Li, Y. Yu, Z.-M. Yang, and L.-L. Tang, "Breast tissue image classification based on semi-supervised locality discriminant projection with kernels," *Journal of Medical Systems*, vol. 36, no. 5, pp. 2779–2786, 2012.

[199] M. Ngadi, A. Amine, and B. Nassih, "A robust approach for mammographic image classification using NSVC algorithm," in *Proceedings of the 1st Mediterranean Conference on Pattern Recognition and Artificial Intelligence (MedPRAI '16)*, vol. Part F126741, pp. 44–49, Algeria, November 2016.