

Research Article

A Diagnostic Model of Breast Cancer Based on Digital Mammogram Images Using Machine Learning Techniques

Farouk A. K. Al-Fahaidy ¹, **Belal Al-Fuhaidi** ², **Ishaq AL-Darouby**¹, **Faheem AL-Abady**¹, **Mohammed AL-Qadry**¹ and **Abdurhman AL-Gamal**¹

¹Faculty of Engineering, Electrical Engineering Department, IBB University, IBB, Yemen

²Faculty of Computing and IT, University of Science and Technology Sana'a, Taiz, Yemen

Correspondence should be addressed to Belal Al-Fuhaidi; belalarh@gmail.com

Received 19 May 2022; Accepted 10 August 2022; Published 20 September 2022

Academic Editor: V. E. Sathishkumar

Copyright © 2022 Farouk A. K. Al-Fahaidy et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Breast cancer disease is one of the most recorded cancers that lead to morbidity and maybe death among women around the world. Recent research statistics have exposed that one from 8 females in the USA and one from 10 females in Europe are contaminated by breast cancer. The challenge with this disease is how to develop a relaxed and fast diagnosing method. One of the attractive ways of early breast cancer diagnosis is based on the mammogram images analysis of the breast using a computer-aided diagnosing (CAD) tool. This paper firstly aimed to propose an efficient method for diagnosing tumors based on mammogram images of breasts using a machine learning approach. Secondly, this paper aimed to the development of a CAD software program for breast cancer diagnosing based on the proposed method in the first step. The followed step-by-step procedure of the proposed method is performed by passing the Mammographic Image Analysis Society (MIAS) through five steps of image preprocessing, image segmentation using seeded region growing (SRG) algorithm, feature extraction using different feature's extraction classes, and important and effectiveness feature selection using the Sequential Forward Selection (SFS) technique, and finally, the Support Vector Machine (SVM) algorithm is used as a binary classifier in two classification levels. The first level classifier is used to categorize the given image as normal or abnormal while the second-level classifier is used for further classifying the abnormal image as either a malignant or benign cancer. The proposed method is studied and investigated in two phases: the training phase and the testing phase, with the MIAS dataset of mammogram images, using 70% and 30% ratios of dataset images for the training and testing sets, respectively. The practical implementation of the proposed method and the graphical user interface (GUI) CAD tool are carried out using MATLAB software. Experimental results of the proposed method have shown that the accuracy of the proposed method reached 100% in classifying images as normal and abnormal mammogram images while the classification accuracy for benign and malignant is equal to 87.1%.

1. Introduction

Breast cancer disease is one of the most recorded cancers which leads to morbidity and maybe death among women around the world. Recent research statistics have exposed that one from 8 females in the USA and one from 10 females in Europe are diseased by breast cancer [1–3]. Hence, breast cancer is the main problem in public health. So, the early detection of this disease is the best stratagem for fighting it. Mammography diagnosis is the most commonly precise tool that can be employed for the early detection of breast cancer

[2, 4, 5]. The preprocessing step of mammogram images is crucial. In research [6], the objective of the Breast Imaging-Reporting and Data System (BI-RADS) of the American College of Radiology (ACR) is to provide a consistent classification system for commentary mammographic breast concentrations. Different research makes the determination either to detect breast lesions using computer-aided detection systems (CADE) or to understand mammograms through Computer-Aided Diagnostic Systems (CADx). These systems are working as a supplement to the radiologists' valuation. In general, there are four steps to be tracked

in the development of a CAD system for diagnosing suspicious regions in mammograms. The first step is the pre-processing for preparing mammograms to be applied in the following steps, without noise. The second one is the identification of regions of interest (ROI) for selecting the desired mammogram information. The third step is the extraction and optimum selection of the features from the identified ROIs. Finally, the classification step of ROIs aims to classify mammograms and decide if it is a normal mammogram or abnormal mammogram which is either a malignant mammogram or a benign mammogram [7, 8].

As it was mentioned above, there are many variant recent works that aimed to propose CAD systems for breast cancer diagnosing, using different feature extraction techniques, such as wavelet processing, statistical methods, and different classification methods like machine learning, neural networks, and deep learning. Recently, AI based on machine learning plays a promising and attractive branch in classification purposes [from tables in related works]. Most of the related works are proposing a classification methodology for benign and malignant while the classification of normal and abnormal is not identified clearly. This work aims to explore a flexible and effective machine learning method for breast cancer diagnosing of normal/abnormal and benign/malignant classifications, with systematic steps and identified methods/techniques. A mammography dataset of the Mammographic Image Analysis Society (MIAS) database [9] is chosen. The proposed CAD system is illustrated in the following steps. The first step, the pre-processing step of the MAIS dataset images, is carried out by applying different image processing techniques such as noise removal, artifacts and labels suppressing, and image enhancement. The 2d median filter is used for noise removal from a mammogram image then, a morphological operation is carried for suppressing artifacts and labels of mammogram images, and finally, the contrast enhancement of mammography images is performed. The second step, the segmentation step, is carried out by taking away the pectoral muscle, using the seeded region growing (SRG) algorithm [10–12]. Subsequently, the identification and extraction of the region of interest (ROI) are done to the result of the segmentation. In the third step, the feature extraction step, several features are extracted from the ROI using different feature extraction classes such as the first-order statistical class, the second-order statistical class based on the gray level cooccurrence matrix (GLCM), the fractal dimension class, the shape class, and the wavelet features class. After that, the step of the optimized feature selection is involved picking the optimum and effective subset of features, which show clear effects on the classification accuracy. This is performed using the Sequential Forward Selection (SFS) technique. Finally, the classification of either normal or cancer by its two different types, benign and malignant, is decided. More details on these steps are presented and discussed in Section 3. The main contributions of this work can be summarized as follows:

- (i) A proposed method based on machine learning for diagnosing breast cancer with mammography

images is introduced and illustrated in systematic steps, studied, and investigated.

- (ii) The MAIS dataset of mammography images is used after carrying out the preprocessing step of the dataset in three substeps, noise removal using 2d median filter, thresholding and contrast enhancement, and pectoral muscle removal using SRG technique.
- (iii) 307 features are extracted using different statistical methods and wavelet feature classes in the suggested method, and only 21-features with an effective impact on the accuracy are chosen during the training phase.
- (iv) The SVM algorithm is used for classification in two levels, one for normal and abnormal classification and the other for benign and malignant classification.
- (v) A CAD system with GUI manipulation is implemented for flexible manipulation of mammography.
- (vi) The accuracy of the proposed method has reached 100% accuracy in the case of normal and abnormal classification of mammography.

The rest of this paper is organized as follows: Section 2 introduces a survey on related works; Section 3 presents and discusses the proposed system. Section 4 demonstrates the practical implementation of the proposed CAD systems and presents and discusses experiments and results. Finally, Section 5 summarizes the conclusions of this work.

2. Literature Review and Related Works

There are numerous approaches that have been anticipated for mammogram diagnosing. In general, they are can be grouped into statistical-based methods [13, 14], wavelets-based methods [15–17], Markovian-based models [18], machine-learning-based methods [19], etc. Numerous investigations have been issued on computer breast cancer diagnosis. In [20], the author presents an overview of the recent advances in the field of CAD breast cancer diagnosis based on mammogram image analysis. In [21], this research presents an outline of procedures that have been suggested in analyzing breast cancer images' histopathology.

For the development of a CAD system for breast cancer, the process of distinguishing the ROIs is an important step that plays a key challenge. Many suggestions have been introduced by researchers for breast tissue/muscle region segmentation according to the density and the texture variances [22]. A proposal provided by [23] uses the Bayesian techniques with a Markovian random field to partition mammogram images into three diverse regions, the pectoral muscle, the fatty, and the fibroglandular regions. Other approaches were using the LBP, the K -means, SVM, and the GLCM algorithms for identifying the ROI regions from mammogram images like in [24–26]. The proposals presented by [2–4] introduce adaptive thresholding methods based on multiresolution for detecting suspicious lesions in mammogram images. For the feature extraction/

TABLE 1: Related works that are carried out for proposing CAD systems for breast cancer.

Ref.	Classifier method	Dataset	Features	Metrics
Jaleel et al. [37]	ANN	MiniMIAS	Discrete wavelet transform and GLCM	Accuracy, sensitivity, specificity
Kaur et al. [38]	SVM, KNN, LDA and DT	MiniMIAS	SURF	Accuracy
Kamil et al. [39]	KNN	MiniMIAS	Gray level Cooccurrence matrix	Accuracy, sensitivity, specificity
Al-Azzam and Shatnawi [40]	Compare supervised learning (SL) with semisupervised learning (SSL) for 9 algorithms	Wisconsin diagnostic breast cancer (WDBC) dataset	30 attributes of 569 patients with 569 instances	Accuracy, sensitivity, specificity
Huang et al. [41]	SVM with three functions and two features: bagging and boosting	Two datasets	First dataset: 11 attributes with 699 instances Second dataset: 117 attributes with 102294 instances	Accuracy, sensitivity, specificity
Asri et al. [42]	SVM, C4.5, NB, k-NN	Wisconsin breast cancer (original) dataset (WBCD)	11 attributes 699 instances	Accuracy, specificity
Rawal [43]	Compare (<i>K</i> -means, EM, PAM, and fuzzy <i>C</i> -means) with SVM and C5.0	Wisconsin prognostic breast cancer dataset	32 attributes with 194 instances	Accuracy
Kourdifi and Bahaj [44]	Random forest, naive Bayes, SVM, KNN	Wisconsin breast cancer dataset	30 attributes with 699 instances	Accuracy, sensitivity, specificity
Cai et al. [45]	Conventional neural network (CNN)	The datasets were collected at two medical institutions	990 images, 540 malignant masses, and 450 benign lesions	Accuracy
Ionescu et al. [46]	CNN	Private dataset	67,520 mammographic images from 16,968 women	Accuracy, sensitivity, specificity
Zebari et al. [47]	Five classifiers with ANN	Mini-MIAS, and other datasets	(i) LBP (ii) FD (iii) Proposed M-FD	Accuracy, sensitivity, specificity

classification step, there are many different types of research suggestions, and the Automatic CAD detection/classification system of suspicious lesions was presented by [27]. In [28], the SVM classification was applied to the development of a classification algorithm for breast masses. The author in [29] presents a different machine learning technique for classifying breast cancer as malignant via cytological imaginings of fine-needle aspiration. The work presented by [30] proposes an automated breast cancer diagnosing system by employing the GVFSnake Segmentation method, the wavelet-based feature extraction, and the fuzzy-based classification. A hybrid optimization algorithm-based feature selection for mammogram images and hybrid transfer learning for detecting the breast masses of mammographs are presented by [31, 32]. The authors in [33] proposed a novel computer-aided diagnosis (CAD) system based on one of the regional deep learning techniques, an ROI-based convolutional neural network for simultaneous detection and classification of breast masses in digital mammograms. In [34, 35], the research is developing a CAD system that employs a temporal analysis for improving radiologists. From the previous studies in the field of breast cancer detection and classification, it is clear to us that the CAD system is an attractive way that could lead to good marks in diagnosing a breast cancer disease. The systematic review in [36] provides a comprehensive description and analysis of

existing CAD systems that make use of machine learning techniques, as well as an assessment of how they now stand in relation to various categorization schemes and mammography image modalities. All CAD phases, including preprocessing, segmentation, feature extraction, feature selection, and classification, were covered in this systematic study. The systematic review outlined suggestions for the next study and identified research gaps.

Table 1 presents some related works and their shortcomings discussed in the following paragraph. In this table, there are different proposed methods with different considerations. Some were working with machine learning like *K*-mean, fuzzy *C*-means, and neural networks (NN) and compared them with SVM, and others using multiclassifiers, artificial neural networks (ANN)s, or convolutional neural networks (CNNs). On another side, there are many different datasets and different feature extraction and selection techniques. Overall, the SVM classifier has approved its practice effectiveness, and the wavelet and GLCM feature extraction classes are performed well. Moreover, there is no clear classification will be identified for normal versus abnormal classification. Hence, we make our decision for carrying out effective preprocessing of the MIAS dataset which is the only available dataset to us in 2018 when we start the work. The preprocessing of the MIAS mammograms is carried out based on the proposed work in [10]. Then, for

feature extraction, we try to use GLCM classes and wavelet classes followed by feature selection using the SFS technique. Finally, based on the literature, the SVM is chosen as a classifier.

3. The Proposed CAD System Methodology

An abstract view of the proposed CAD system processes is shown in Figure 1. The flow of processes of the suggested CAD system is starting by preprocessing the mammogram images for removing noise. The next step, image segmentation, is carried out for detecting and identifying the ROIs of mammogram images. Then, the ROIs are used in the features extracting step. After that, the Sequential Forward Selection (SFS) technique is used for the important and relevant feature selection. The final step, the classification, is applied to mammogram's features to classify mammograms as either normal or abnormal, with further classification of abnormal images as benign or malignant. A detailed description of these steps is presented in the followed subsections.

3.1. Digital Mammography Image Dataset. As an initial step, before the processes of the proposed CAD system for breast cancer diagnosing system, data collection will be obtained.

3.1.1. Data Collection. We obtain a well-known digital mammogram dataset from a known data acquisition society called the mini-MIAS database. The obtained images include left and right breast images of breasts that are fatty, fatty-glandular, and dense-glandular. The three basic cases of the obtained mammogram images are malignant, benign, and normal that are each further categorized into five groups as follows:

- (1) Constrained masses
- (2) Spoilage-like masses
- (3) Irregular masses
- (4) Masses with deformed architecture
- (5) Asymmetries in the masses

A well-known available mammogram images database which represents 322 digital mammogram images of 161 pairs at a resolution of 50μ as a portable gray map style with associated truth data, from the MIAS, UK [9], is selected. The MIAS data is used to feed as standard inputs to the proposed CAD system in this work. Images are associated with annotation labels.

3.2. The First Step: The Preprocessing of Database. Naturally, different medical images, such as mammogram images, are hard to understand or interpret; therefore, preprocessing is required to enhance the quality of images by removing noise and to make better results in the segmentation step. Image preprocessing includes noise removal, artifact suppression, and background separation [10].

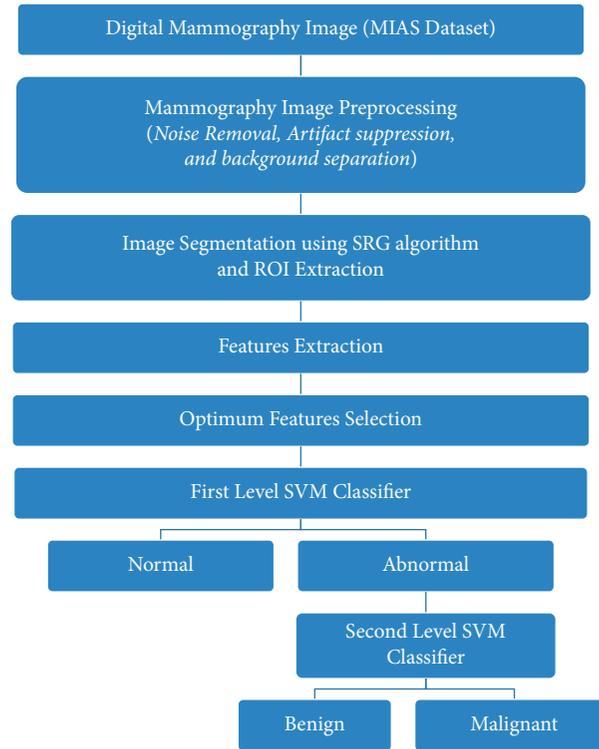


FIGURE 1: A flow diagram of the proposed CAD system methodology.

3.2.1. Noise Removal. The majority of obtained mammography images contain digitization noises like straight lines, which are filtered using a two-dimensional (2D) median filtering technique in a 3-by-3 neighborhood connection. Each output pixel includes the median value for the 3-by-3 area surrounding the corresponding input pixel for removing artifacts, labels, and unwanted image borders by using morphological operations. However, the images' edges are changed to zeros. A mammography image is shown in Figure 2(a) with the digitization noise present, and the same image is shown in Figure 2(b) with the noise removed.

3.2.2. Artifact Suppression and Background Separation. Using threshold and morphological techniques, shadow artifacts in the mammography images such as wedges and labels are eliminated. Figure 2(a) displays a mammography image with a shadow artifact that was taken from the MIAS mammogram database. A global threshold with a value of $= 18$ normalized value is discovered to be the most suitable threshold for converting the grayscale images into binary [0, 1] format through manual inspection of all mammography images acquired. The binary mammography pictures are subjected to morphological operations such as dilation, erosion, opening, and closing after the grayscale mammogram images are converted to binary, as shown in Figure 2(c) for the image in Figure 2(a). The breast profile region is likewise separated from the background during artifact, wedge, and label suppression (see Figure 2(d)).

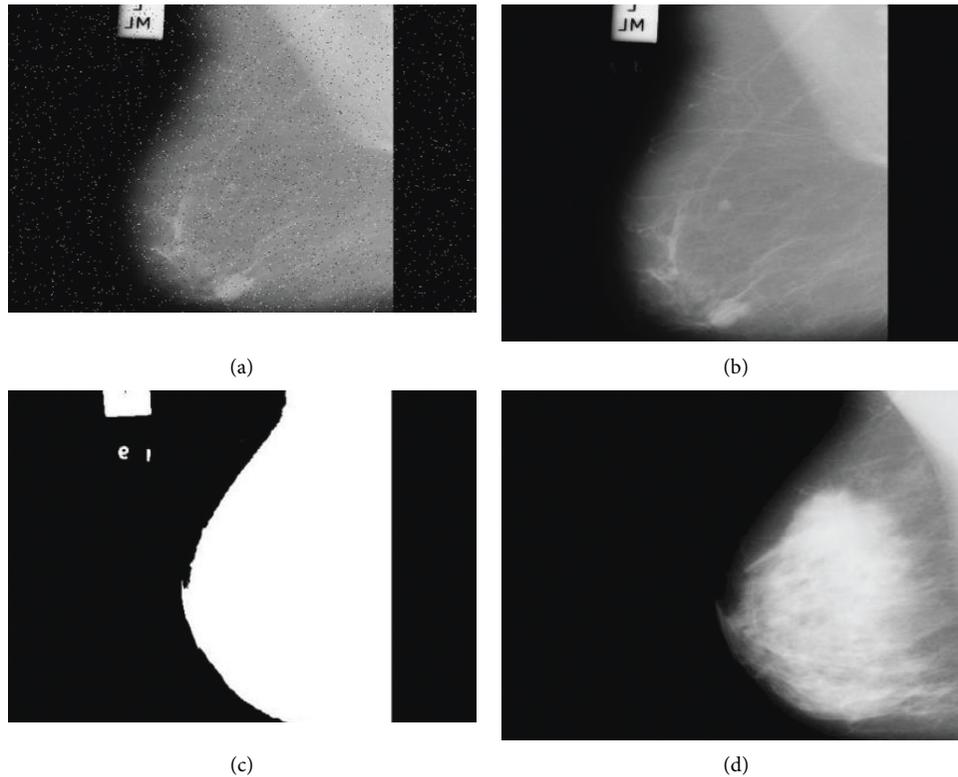


FIGURE 2: The preprocessing experiments of the input image. (a) The original image with noise. (b) Image after removing the noise. (c) The binary image after background separation. (d) The output image, the breast profile only.

Figure 3 presents a sample mammogram image with different types of elements before applying the preprocessing and segmentation steps.

3.3. The Image Segmentation Step. The main objective of image segmentation is to identify and extract the ROI, i.e., the pectoral muscle, which has similar characteristics that may match the tumor in the breast profile if exist.

Initially, for obtaining the best outcomes of segmentation, we do the following steps:

- (1) Prior to completing seeded region growing (SRG), it is necessary to determine the breast orientation in each mammography image. The binary image in Figure 2(c) is used to determine the breast profile orientation (left or right) using an automated approach. The binary image is chopped from top to bottom and from left to right so that the breast profile hits the image's four boundaries (top, left, right, and bottom). In the cropped binary images, the sum of the first and final five columns of binary values is then determined. The breast profiles are oriented by simply comparing the sums of the first and final five columns; if the first sum is bigger than the last, the breast is right-oriented; otherwise, it is left-oriented.
- (2) The image contrast of the breast profile images is enhanced using image adjusting and stretching limits techniques in MATLAB.

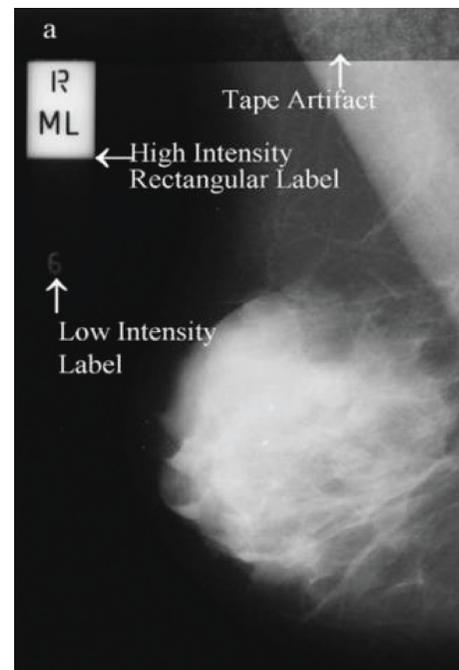


FIGURE 3: Mammogram image with various elements.

- (3) Then, the image segmentation is carried out via the SRG algorithm for extracting the pectoral muscle from a processed mammogram image. For more details on the SRG algorithm, see [11, 12].

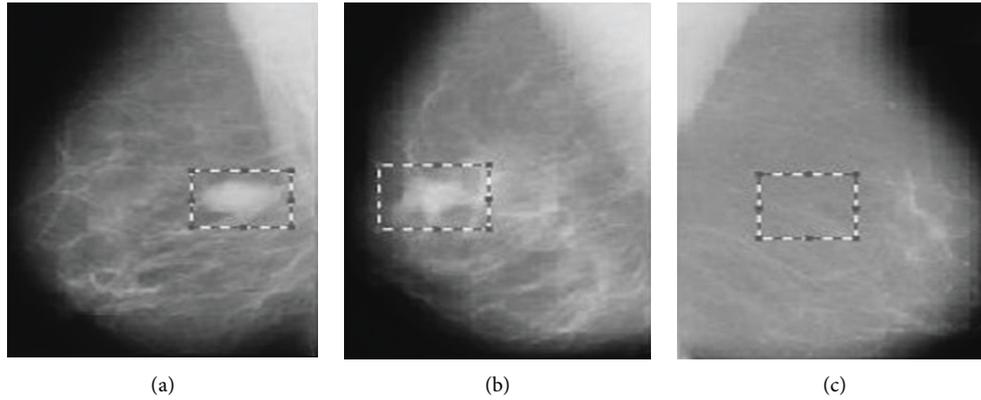


FIGURE 4: An example of ROIs within Mammogram images. (a) Benign ROI. (b) Malignant ROI. (c) Normal ROI.

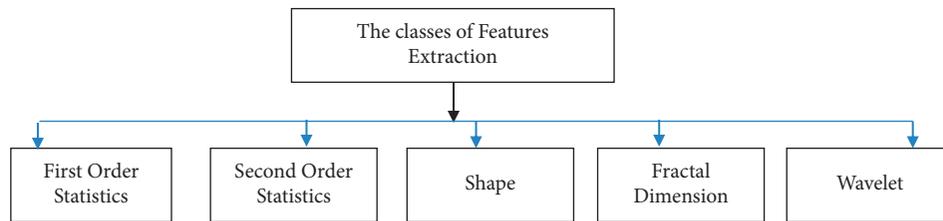


FIGURE 5: Feature extraction main classes.

3.3.1. The Region of Interest (ROI) Selection. Features necessity is to be calculated as of abnormal region, benign and malignant, in breast profile without including all other insignificant fragments of the breast tissue. So, features could not be directly calculated from the segmented mammogram images. That is due to the fact that the segmented image willpower biases the exposure results. For such cases, the Ground Truth data assimilated along with the mammography's dataset information are used for identifying as well as for extracting the ROIs [9]. According to the size of ROI, a suitable ROI window size that is adequate for encircling the majority of benign and malignant abnormalities is selected to be 64×64 pixels [23, 24]. So, the proposed CAD system uses an ROI block of size 64×64 pixels. Figure 4 shows examples of ROIs within 3-different mammogram images, Figure 4(a) is an example of benign ROI, Figure 4(b) is an example of malignant ROI, and Figure 4(c) is an example of normal ROI.

3.4. The Feature Extraction Step. Since the conventional images are complicated to manipulate and are highly textured, they led to difficult interpretation. So, the digital mammogram images are used in the feature extracting phase of the proposed system based on their ability for visualizing masses. This is because the spatial resolution of digital mammogram images resulting from an X-ray is in the order of limited microns. Consequently, it is necessary to enhance the performance of the diagnosis in terms of accuracy and reliability by extracting features from digital mammograms. Figure 5 shows five classes of feature extraction methods that are applied in this work.

TABLE 2: List of eighteen first-order statistical features.

Mean	Maximum gray level
Standard deviation	Minimum gray level
Kurtosis	Entropy
Skewness	Second moment
Variance	9-percentiles

An explanation of these five main classes of feature extraction is presented in the following subsections.

3.4.1. The Features of First-Order Statistics Class. The first-order statistics feature class offers variant statistical properties representing the image's intensity histogram. The first-order statistic depends on the intensity values of each pixel only. Table 2 lists 18-features representing the first-order textural features that could be extracted from ROI [48, 49].

3.4.2. The Features of Second-Order Statistics Class. The most commonly used second-order statistics feature is the GLCM feature, which can be applied in the extraction of 2nd order texture information on images based on well-known features having a strong statistical tool. Table 3 presents a list of GLCM features [50–52].

3.4.3. The Features of the Shape Class. To notice the ROI's shape of images, the shape features' class could be employed. Many shape features can be extracted from mammogram images; the quantitative analysis of these shape's properties permits us to distinguish between malignant lockups from those of benign as well as the normal cells. Eight shape

TABLE 3: List of the second-order statistics features.

Energy	Shade	Information correlation 2
Contrast	Prominence	Sum of squares
Homogeneity	Max of cooccurrence	Sum average
Correlation	First-order difference moment	Sum entropy
Entropy	Second-order difference moment	Difference entropy

features could be derived from each ROI, the spread-ness, and the seven invariant moments [48, 53].

3.4.4. The Features of the Fractal Dimension Class. The fractal is an asymmetrical geometric entity with infinite nested structures on its whole scale. The most significant properties of fractals are chaos, self-similarity, and fractal dimension. The fractal dimension class provides a calculable measure of the similarity-on-self and the scaling features. The two ways which are used to estimate the fractal dimension features and used by the proposed system in this work are as follows:

The piecewise triangular prism surface area (PTPSA) technique,

The piecewise modified box-counting (PMBC) technique.

The exponent of the number of similarity-on-self fragments, N , to the magnification factor, $1/R$ into which a figure may be fragmented defines the fractal dimension [54], as listed by

$$FD = \frac{\ln(N)}{\ln(1/R)}. \quad (1)$$

3.4.5. The Wavelet Features' Class. One of the most commonly used signal/image processing/representation techniques is wavelet transformation. Typically, wavelet transformation of images results in two types of decomposition, the detail, and the approximation coefficients. The detailed coefficients represent high-frequency components of the input image; the approximations represent low-frequency components of the input image. Wavelet transform is good in extracting the characteristics' vectors. In this numerical signature, the vector of characteristics represents masses and calcifications through the smaller-scaled image with a minimum amount of values; this depends on the resolution of the input image [55, 56]. In this work, the wavelet transform is used for detecting and segmenting masses and calcifications, by evaluating the practice of the Haar and the Daubechies wavelet types for breast lesions, mass, and calcification and characterization. These two wavelet transforms can characterize the difference between masses and calcifications over the gray levels. Figure 6 demonstrates the wavelet processing of the image as a low-frequency filter is used to pass low-frequency components resulting in approximation coefficients of the input image while a high-pass filter that passes high-frequency components of the image results in detailed components.

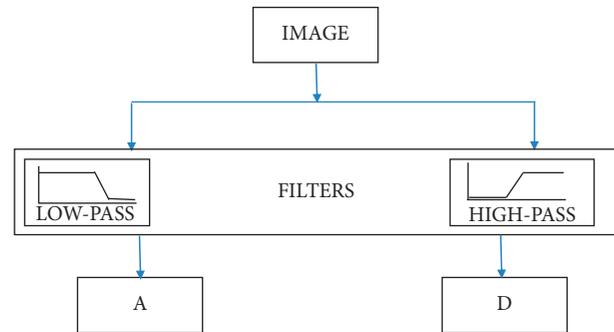


FIGURE 6: The discrete wavelet decomposition processing of an image into approximations and details components.

An example of applying the Haar wavelet transform is presented in Figure 7, with two levels of decomposition. The approximation coefficient informs nuance while, the detailed coefficients, in horizontal, diagonal, and vertical provide the image its identity.

3.5. The Important Feature Selection Step. The critical role in providing the best performance of the classifier of the image is how accurately the features are extracted and selected by the classifier. The optimum selection of the effective features from the whole feature set provides the best accuracy of the classifier. Roughly, there are three different categories of feature selection methods which are as follows: the filtering methods, the Wrapper methods, and the embedded methods. In this work, the Sequential Forward Selection (SFS) technique, which belongs to the Wrapper methods, is used for the selection of the important features from the extracted features [57, 58]. In the SFS selection method, an empty set of features is growing sequentially, by adding variant features to the feature set, training the model with these added features, and dropping out features with the lowest effects, until the addition of extra features does not reduce the criterion.

3.6. The Classification Step. The final stage of the CAD system is the classification which is used to categorize breast cancer to be either normal or abnormal as malignant or benign. It aids in predicting features using the Support Vector Machine (SVM) algorithm. SVM was established by the machine learning community [59]. The SVM is a classifier defined by a separation of hyperplanes. The hyperplane is defined by the distance from such hyperplane to the nearest data points on each side, termed maximal support vectors [60]. It could be extended to nonlinearly separable data with the help of kernel function application on the data

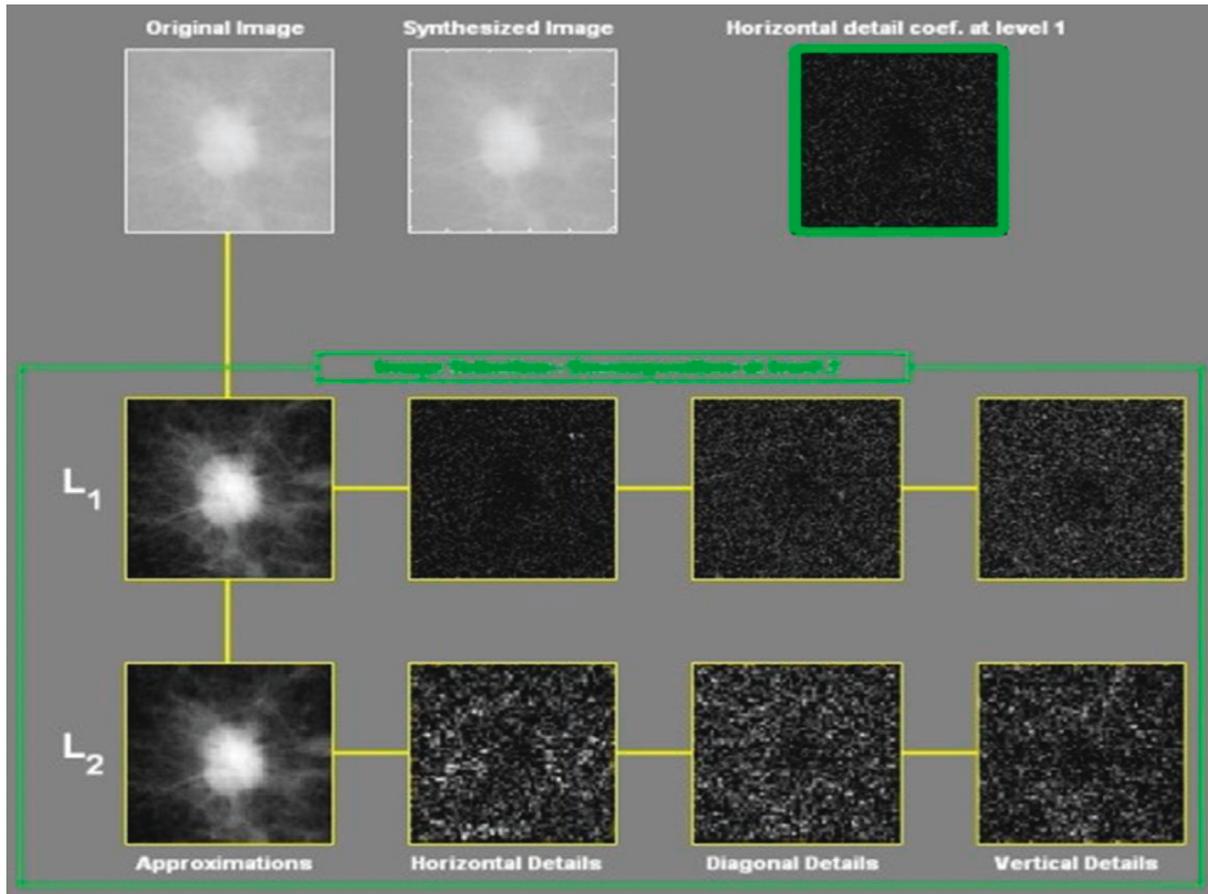


FIGURE 7: A view of Haar wavelet features.

to make them linearly separable [61]. An approach based on wavelet SVM was discussed by [62]. More details on SVM and its application to the diagnosing of breast tumor was discussed in [63, 64].

4. Computer Experiment Study and Investigation

This section presents the experimental results as well as the implementation of the proposed computer-aided detection and diagnosis (CADD) system. Different experiments have been carried out in the sequential stages followed by the CADD system. The next subsections are providing the experimental results of the preprocessing, image segmentation, feature extraction and selection, and classification stages. Then, the implementation of the proposed CADD system in the working environment, the MATLAB software, is presented, discussed, and tested.

4.1. The Preprocessing. The preprocessing is carried out by applying a mammogram image to be passed through the nonlinear median filter for noise removal; then, the filtered binary image is separated from the background. Finally, the only breast profile is taken as an output from this phase. Figure 2 shows the outputs of these consecutive experiments that are carried to the input processed mammogram image.

The input image is shown in Figure 2(a), and the filtered image after noise removal is presented in Figure 2(b), which shows a significant improvement against the input image in its clearance. Figure 2(c) demonstrates the image's background separation. Finally, the breast profile of the filtered image after the separation of background is presented in Figure 2(d).

The output of this stage, the breast profile, is applied as input to the segmentation step for extracting the ROI.

4.2. The Image Segmentation. The objective of the segmentation phase is to identify and extract the pectoral muscle, i.e., the ROI. Initially, in the output of the preprocessing step, the breast profile of the input image's contrast is enhanced for obtaining the finest results from segmentation. Then, the image segmentation is carried out by applying the SRG algorithm for extracting the pectoral muscle from a processed mammogram image. Figure 8(a) shows the input image after contrast enhancement with identified seed point while Figure 8(b) presents the pectoral muscle extraction.

4.3. The Feature Extraction. In this phase, the pectoral muscle is fed as input to the five classes of feature extraction presented in Figure 5. The output of feature extraction is 307 features that are extracted from each ROI. These features are organized as follows:

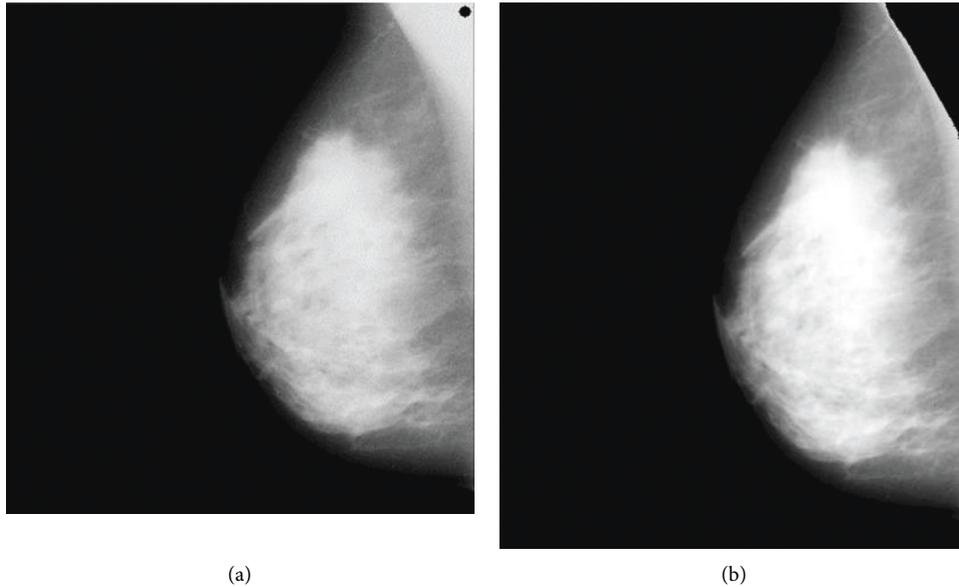


FIGURE 8: Image segmentation result. (a) The preprocessed image after enhancement with identified seed point. (b) Image after segmentation (pectoral muscle extraction).

	1	2	3	4	5	6	7	8
1	231.8472	8.8965	2530	1.4493e+20	252	4.9503	53	79.1276
2	234.9563	13.2697	2179	3.0884e+19	255	5.3643	72	176.0418
3	169.3303	4.3119	3397	7.4730e+20	184	3.8515	27	18.5878
4	196.6301	14.1488	2014	1.1721e+19	236	5.4580	86	200.1402
5	190.1196	15.3248	3046	7.4431e+18	237	5.5550	82	234.7918
6	150.1885	8.3047	3759	3.3610e+19	178	4.9069	48	68.9524
7	206.0103	7.9245	2772	1.4352e+20	234	4.7187	56	62.7822
8	220.5400	7.4874	3963	2.3651e+20	242	4.6869	41	56.0477
9	215.2615	22.5495	1866	2.6094e+18	255	5.8127	86	508.3546
10	200.3257	49.7778	1547	8.2420e+16	255	6.3525	175	2.4772e+03
11	189.5708	8.1656	3537	9.1277e+19	216	4.8313	48	66.6603
12	247.6895	7.8891	3540	3.0532e+20	255	3.8133	36	62.2224
13	235.0337	15.2536	2118	1.7711e+19	255	5.0656	74	232.6165
14	129.8301	17.8975	1907	8.7009e+17	204	5.6045	102	320.2411
15	130.3000	5.8191	2416	7.8992e+19	150	4.2554	38	33.8533
16	237.0994	8.8719	2527	1.6028e+20	255	4.8034	45	78.6923
17	231.6963	16.1098	2167	1.3445e+19	255	5.4035	75	259.4625
18	188.0093	29.4949	1578	5.1875e+17	224	5.7950	174	869.7376
19	78.7691	50.3950	1660	1.9769e+15	197	6.0096	197	2.5390e+02

FIGURE 9: A snapshot of a portion of statistical features.

- (i) The first-order statistics features that are extracted are 10 features
- (ii) The second-order statistics features that are extracted are 62 features
- (iii) The shape features that are extracted are 9 features
- (iv) The fractal dimension features that are extracted are 2 features
- (v) Finally, the wavelet features that are extracted are 224 features

From these features, the optimum and effective small number of features are selected by the next step, the feature selection, to be used in the final stage, the classification stage.

A snapshot of a sample of statistical features which represent some different first- and second-order statistical features matrix values is shown in Figure 9. Figure 10 shows a portion of wavelet features matrix values extracted by feature experiments. It should be noted that each row in both of the previous figures numbered Figures 9 and 10 represents a sample image and the columns contain the values of the extracted features.

	1	2	3	4	5	6	7	8
1	13.5120	13.5245	13.5153	13.5371	13.4787	13.6437	13.6231	13.6524
2	12.6379	12.6423	12.6366	12.6478	12.6213	12.6943	13.1236	13.3358
3	7.2307	7.2596	7.2445	7.3032	7.1379	7.5829	8.4331	9.1150
4	9.7839	9.8101	9.7897	9.8535	9.6671	10.2080	10.3928	10.3888
5	9.9901	10.0082	9.9936	10.0393	9.8989	10.3170	11.0458	11.4576
6	11.4505	11.4707	11.4609	11.5024	11.3697	11.7504	12.0309	12.0832
7	12.4711	12.4738	12.4688	12.4840	12.4414	12.5705	12.6908	12.7130
8	9.6121	9.6179	9.6065	9.6503	9.4928	10.0185	10.9634	11.4284
9	12.0247	12.0371	12.0262	12.0551	11.9718	12.2204	12.4656	12.8192
10	10.9616	10.9907	10.9757	11.0219	10.9014	11.1977	11.8186	12.4297
11	11.7057	11.7023	11.7018	11.6971	11.7160	11.6725	11.3352	10.8479
12	7.6091	7.6270	7.6090	7.6666	7.4864	8.0406	8.6156	9.1587
13	10.9141	10.9207	10.9260	10.9301	10.9001	10.9743	11.1785	11.3591
14	7.9502	7.9581	7.9607	7.9870	7.8855	8.1524	8.7504	10.2642
15	10.6643	10.6299	10.6496	10.6148	10.6679	10.6267	10.4085	9.9819
16	8.6710	8.7145	8.6901	8.7931	8.4685	9.4074	10.8970	11.6488
17	11.7907	11.8084	11.7946	11.8442	11.6900	12.1436	12.4805	12.5864
18	11.5412	11.5631	11.5457	11.5756	11.5255	11.6172	11.8369	11.7105
19	10.7841	10.7861	10.7822	10.7831	10.7604	10.7731	10.7758	11.4822

FIGURE 10: A snapshot of a portion of wavelet features.

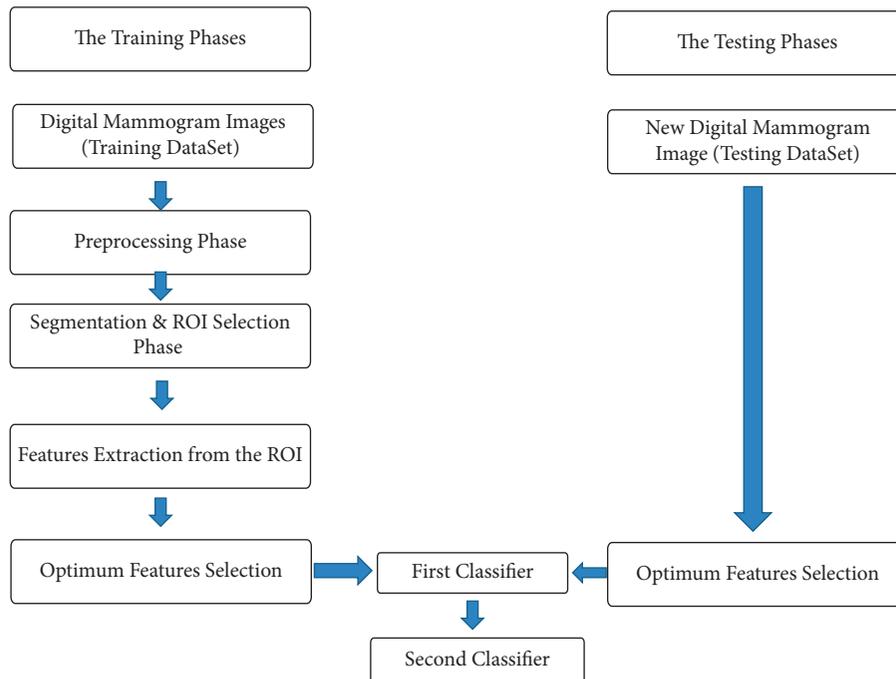


FIGURE 11: The training and testing phases following the diagram.

4.4. The Important Feature Selection. As shown in the previous Section 4.3, there are a total number of 307 features that are extracted based on different types of features which is a large number of features. This phase aims to optimize the number of features by selecting the optimum number of effective features. This is performed using the SFS technique with exhaustive experiments carried out and examined in the two steps of the training and the testing phases using samples around 70% and 30%, respectively. More

explanation on learning and testing samples ratios and results is shown in the next subsection, the classification. Exhaustive experiments are carried out in training and testing phases for selecting the effective features based on the steps illustrated in Figure 11. As a result of these experiments, 5 features are selected from the total of 83 statistical features, and 16 features are selected from the total of 224 wavelet features. The selected features show their importance thru the training phase; i.e., there are a total of 21

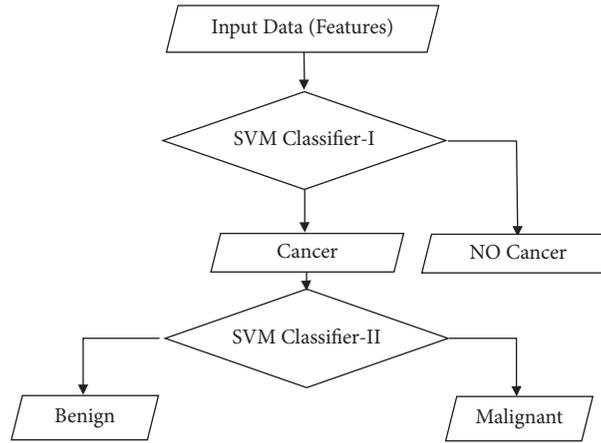


FIGURE 12: The flow of the two SVM classifiers.

TABLE 4: The training and testing ratios distribution of samples from the MIAS database.

The class type	Number of samples	Training set (around 70% of samples)		Testing set (30% of samples)
Normal (no cancer)	208	157		51
Abnormal (malignant)	49	37		12
Abnormal (benign)	67	48		19
Total number	324	242		82

effective features that are selected during the training phase. These 21 features are then used for the final version of the proposed CADD system in a testing phase.

4.5. The Classification. A well-known classifier algorithm, the SVM, is used for the classification stage of the proposed CADD system. Since the SVM is a binary classifier, in this work, two SVM classifiers are used to avoid this problem as indicated in Figure 12.

In the operation of these two SVM classifiers as illustrated in Figure 12, the first SVM classifier is used to identify if the entered data, the features, indicate cancer or no cancer, i.e., abnormal class or normal class. Then, based on the resulting class from the first classifier, the second SVM classifier may be activated or inactivated. Intuitively, if the first classifier indicates a normal case, the two outputs of the second classifier will be 0. Otherwise, the second SVM classifier is activated to decide on the type of cancer, either a malignant class or a benign class.

4.5.1. Performance Evaluation of the SVM Classifier. To implement the SVM classifier, the selected features from the mammogram image are separated into two distinct sets, the learning set, and the testing set. Table 4 shows the total number of samples, 324 samples which are separated into two main groups, around 70% of total samples are used for the learning/training set and the remainder, around 30% of samples, are used for the testing set. From these two sets, there are 208 samples of normal classes and 116 of abnormal classes with 49 samples belonging to the cancer class and 67 samples belonging to the benign cancer class.

The training accuracy of the SVM classification machine could be estimated by regulating the error parameter. Our proposed system uses the RBF (Gaussian) with nonlinear SVM classification. The RBF kernel's parameter is employed to control the width of the Gaussian and needs to be optimized for the SVM hyperparameter. Therefore, the determination of two SVM hyperparameters is needed for constructing an optimum classifier that balances its generalization and memorization capabilities.

Now, the results of testing of the two working SVM classifiers are obtained and presented for dialogue in Tables 5 and 6, of the two binary confusion matrices:

Upon the two confusion matrices listed in previous tables, the performance computation could be done by calculating the *Sensitivity*, *Specificity*, and *Accuracy* values of the two SVM classifiers according to the following equations:

$$\begin{aligned}
 \text{Sensitivity} &= \frac{\text{Positive correctly classified}}{\text{Total Positive}} = \frac{TP}{TP + FN} \\
 \text{Specificity} &= \frac{\text{Negatives correctly classified}}{\text{Total negatives}} = \frac{TN}{TN + FP} \\
 \text{Accuracy} &= \frac{\text{Instances correctly classified}}{\text{Total instances}} = \frac{TP + TN}{TP + TN + FP + FN}
 \end{aligned} \tag{2}$$

The results of the first SVM classifier, in the testing of normal and abnormal classes, show that the classifier's *Sensitivity*, *Specificity*, and *Accuracy* are all 100% which is a great performance while the second SVM classifier, in the testing of malignant and benign classes, shows that the classifier's *Sensitivity* is equal to 90%, the *Specificity* is 85.714%, and the *Accuracy* is equal to 87.1%; this is due to the use of few numbers of images for diagnosing malignant and benign cancer types. However, the second SVM classifier is

TABLE 5: The confusion matrix of the first SVM classifier, in the testing of normal and abnormal.

The class type	Correct classified	False classified	Total number
Normal (positive class)	TPs = 51	FPS = 0	51
Abnormal (negative class)	TNs = 31	FNs = 0	31
			82

TABLE 6: The confusion matrix of the second SVM classifier, in the testing of malignant and benign.

The class type	Correct classified	False classified	Total number
Positive class (malignant)	TPs = 9	FPS = 3	12
Negative class (benign)	TNs = 18	FNs = 1	19
			31

providing significantly high performance. By looking at the results presented in Table 6, it is noticed that the malignant class provides a bad performance compared to all other classes; this is maybe due to the malignant class being trained and tested with the smallest number of samples of 37 samples in training and 12 samples in testing. So, it can be noted that the performance of the second classifier could be improved by applying an extended dataset of mammogram images which may be available soon.

The suggested system has shown its capability for diagnosing breast cancers with the MIAS dataset. However, this dataset consists of a few hundred images. Therefore, it is noticeable for improving the performance of the suggested system in terms of its accuracy. For dataset size, firstly, we increase the number of available MIAS database images, by applying an augmentation process to images. This is carried out via constructing many forms of images from a given image by applying some augmentation operations, such as different angel rotations and flipping. Secondly, we use other datasets with a larger size and even use many different datasets available on the Internet. Another way for the system model is that one can use a deep learning model based on convolution neural networks or any pretrained convolution neural networks.

4.5.2. Comparative Study and Limitations. This section compares the implemented algorithm in this work with related works to assess how well it performs. The comparison is based on the accuracy metric. The studies that accessed the MIAS database and different databases are used with SVM classifiers and different classifiers.

According to the related works in Table7, especially those that used few images of the MIAS dataset with the GLCM feature extraction, the proposed method gives a good accuracy compared to the other methods that used KNN [39] and SVM [39] classifiers. Other methods used a large number of images in many datasets along with MIAS for training the model. The authors in [40, 41, 45, 47] trained the proposed methods with different classes of features and different classifiers algorithms, and they outperform our proposed method in terms of accuracy metrics, due to the large numbers of images and different datasets. Our proposed method is characterized by three properties. The first property, the use of two classifiers, has trained the suggested

method for classifying cancer as normal/abnormal and benign/malignant while they are only trained and classified cancer as benign versus malignant. The second property is that our suggested method has reached 100% accuracy in classifying normal and abnormal mammograms. And the third property is characterized by the implementation of a computer program with GUI. However, the second-level classifier of benign versus normal of the suggested method reached 87.1% in terms of accuracy. This is due to the minimum number of images that are used. Henceforward, for improving the accuracy of the second classifier, one can either retrain the suggested model with more mammograms either by using of additional dataset or by extending the existed images using augmentation methods to these mammograms.

4.6. The Implementation of the Proposed Computer-Aided Detection and Diagnosis (CADD) System. This work is finalized by implementing an application of the proposed CADD system using MATLAB software environment, version R2013a, on the windows 8 operating system, by employing MATLAB's image processing tools and statistical tools. All experiments are implemented on a Dell laptop of Intel Core i7 with the processing power of 2.4 GHz CPU with 12 GB RAM. Figure 13 illustrates the graphical user interface working environment of the proposed CADD system.

Figure 14 shows the relationship between different files used in the development of the proposed CADD system application. The dashed spherical shape representation means there are additional functions to this component/file code. These files/components and their operations are summarized as follows:

- (i) *Image reading.m*: this file allows the user to enter the mammographic image from a specific file directory to be diagnosed by the proposed CADD system.
- (ii) *Image preprocessing.m*: This file improves the quality of the input image and removes the noise. It includes the following MATLAB files:
 - (i) *Binary.m* which eliminates the noise and converts the image into a binary image.
 - (ii) *Breast.m* which removes the labels and identifies the orientation of the image.

TABLE 7: The comparative study with related works.

Related work	Classifier method	Dataset	Features	Accuracy of normal/ abnormal classifier	Accuracy of benign/ malignant classifier	CAD tool based-GUI
Kamil Jassam [39]	KNN	MIAS	Gray level cooccurrence matrix (GLCM)	None	86.1%	No
Al-Azzam and Shatnawi [40]	Compare supervised learning (SL) with semisupervised learning (SSL) for 9 algorithms	Wisconsin diagnostic breast cancer (WDBC) dataset	30 attributes of 569 patients with 569 instances	None	K-NN (SL = 98% & SSL = 97%) and logistics regression (SL = 97% & SSL = 98%)	None
Huang et al. [41]	SVM with three functions and two features: bagging and boosting	Two datasets	First dataset: 11 attributes with 699 instances Second dataset: 117 attributes with 102294 instances	None	96.85%, 95%	None
Cai et al. [45]	Conventional neural network (CNN)	The datasets were collected at two medical institutions	990 images, 540 malignant masses, and 450 benign lesions	None	87.68%	No
Zebari et al. [47]	Five classifiers with ANN	Mini-MIAS, and other datasets	(i) LBP (ii) FD (iii) Propose M-FD	None	Train, test 79.11, 83.61 80.06, 89.55 96.2, 100	No
Kamil and Jassam [39]	SVM	MIAS	GLCM	None	83.5%	
Our work	Two-level SVM classifiers	MIAS dataset	Statistical classes, wavelet class, and SFS technique: 21 attributes with 307 instances	100%	87.1%	Yes

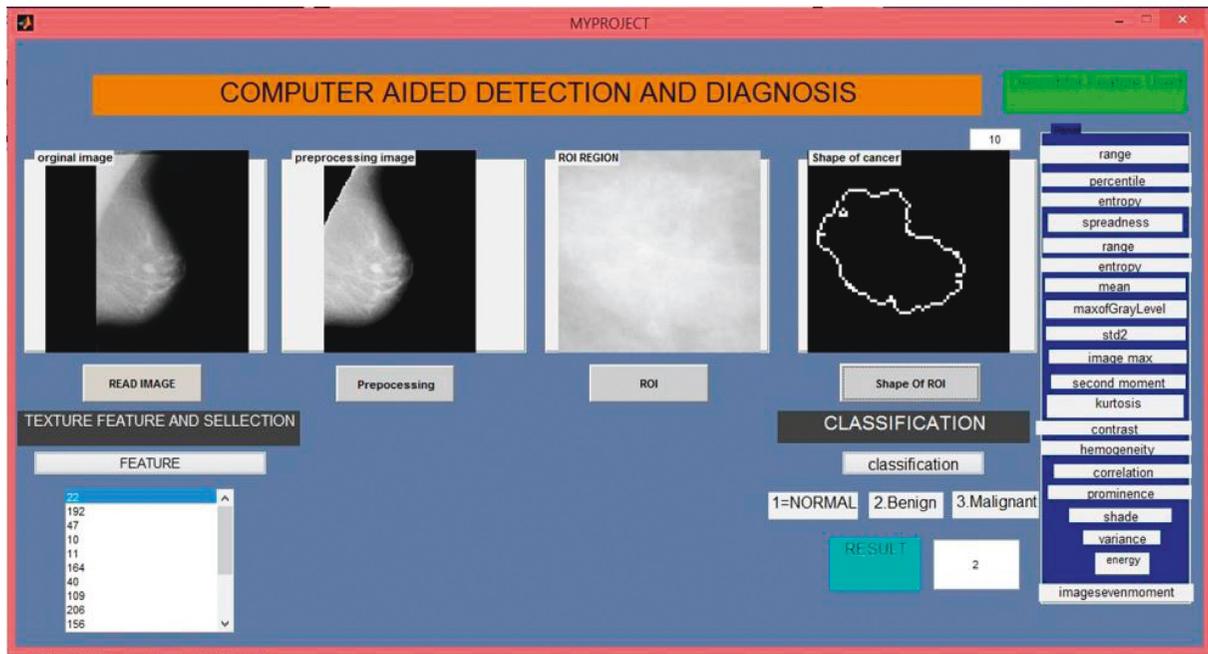


FIGURE 13: The proposed CADD system application.

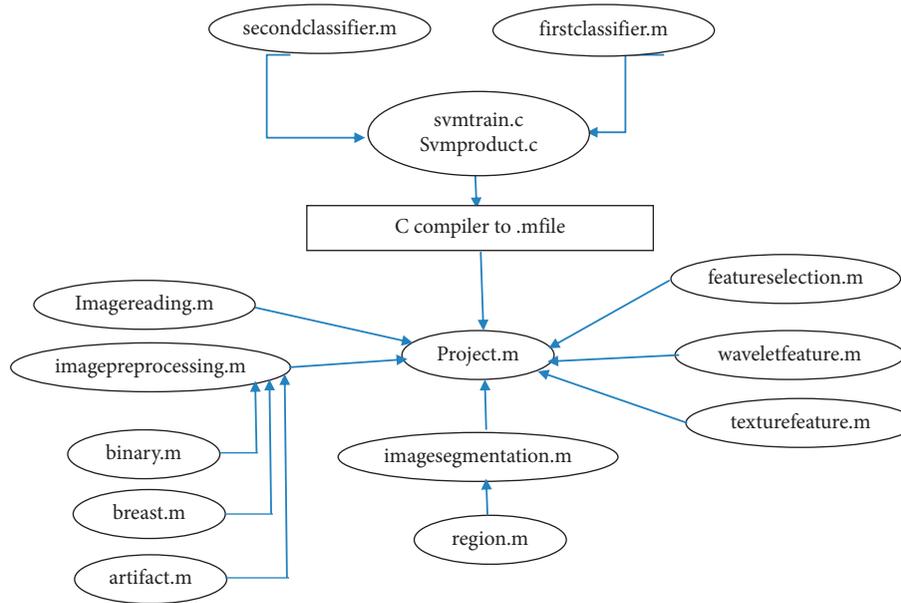


FIGURE 14: The relation between different MATLAB files used in the development of the proposed CADD system Application.

- (iii) *Artifact.m* which removes the artifacts from the image and separates the background from the breast profile.
- (iii) *Image segmentation.m*: this file implements the contrast enhancement method and removes the pectoral muscle from the preprocessed image. It includes the *region.m* file which implements the pectoral muscle removal by using the SRG algorithm (implemented in the *region-growing.m* file) and enhances the image after pectoral muscle removal.
- (iv) *Texture feature.m*: this file extracts 83 texture features from the segmented image.
- (v) *Wavelet-feature.m*: this file extracts 224 wavelet features from the segmented image.
- (vi) *Feature-selection.m*: it selects the optimum texture and wavelet features for providing better classification accuracy. It employs the SFS algorithm which selects 5 of 83 texture features and 16 from 224 wavelet features.
- (vii) *First-classifier.m*: it implements the first SVM classifier by using the selected 5 texture features to classify cases as normal and cancer cases. It includes the following files:
 - (i) *svmtrain.c* which trains the first SVM classifier.
 - (ii) *svmpredict.c* which tests the first trained SVM model.
 - (iii) *svmtrain.c* and *svmpredict.c* which are implemented by using the visual studio compiler.
- (viii) *secondclassifier.m*: it implements the second SVM classifier by using the selected 16 wavelet features to classify cases as benign and malignant cases. This file includes the following files:
 - (i) *svmtrain.c* which trains the second SVM classifier.
 - (ii) *svmpredict.c* which tests the second trained SVM model.
 - (iii) *project.m*: which contains the graphical user interface which gathers all the above files into this file.

5. Conclusions

This work has proposed a CAD system for breast cancer identification and diagnosis from mammography images. The proposed method is passed throughout followed and ordered systematic steps for manipulating the mammography images to identify the ROI, feature extraction and optimization, and then for breast cancer classification using a machine learning algorithm, the SVM algorithm. The experimental results of the proposed method have shown its effectiveness, especially in the case of normal and abnormal tumor classifications which reach 100% accuracy. After the investigation of the proposed method with training and testing phases, a CAD software system with GUI is implemented using the proposed method. This CAD tool can play as an efficient tool for early identification and diagnosis of breast cancer from mammogram images. However, the second-level classifier accuracy reached 87% in classifying the cancer class as either benign or malignant. This is due to the small size of the available MAIS dataset

images, especially mammogram images with tumors. So, we suggest different future works be considered. One of the future works to be suggested is to extend the size of the MAIS dataset using augmentation. Another suggestion is to use another dataset with a larger size. Moreover, researchers may have to use a suitable deep learning machine like the conventional neural networks.

Data Availability

The mini MIAS images data set used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] A. Elmoufidi, K. El Fahssi, S. Jai-Andaloussi, and A. Sekkaki, "Automatically density based breast segmentation for mammograms by using dynamic k-means algorithm and seed based region growing," in *Proceedings of the I2MTC 2015-International Instrumentation and Measurement Technology Conference*, Pisa, Italy, May 2015.
- [2] K. Hu, X. Gao, and F Li, "Detection of suspicious lesions by adaptive thresholding based on multiresolution analysis in mammograms," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 2, pp. 462–472, 2011.
- [3] J. Anitha, J. Dinesh Peter, and S. Immanuel Alex Pandian, "A dual stage adaptive thresholding (DuSAT) for automatic mass detection in mammograms," *Computer Methods and Programs in Biomedicine*, vol. 138, pp. 93–104, 2017, p.
- [4] B. C. Patel, G. R. Sinha, and D. Soni, "Detection of masses in mammographic breast cancer images using modified histogram based adaptive thresholding (MHAT) method," *International Journal of Biomedical Engineering and Technology*, vol. 29, pp. 134–154, 2019, p.
- [5] A. Ferrero, S. Simona, M. Arianna, R. Giulia, and S. Marcello, "Uncertainty evaluation in a fuzzy classifier for microcalcifications in digital mammography," in *Proceedings of the I2MTC 2010—International Instrumentation and Measurement Technology Conference Austin*, pp. 3–6, Austin, TX, USA, May 2010.
- [6] American College of Radiology, *American College of Radiology Breast Imaging Reporting and Data System (BIRADS)*, American College of Radiology, Reston, VA, USA, 2003.
- [7] A. Jalalian, S. B. Mashohor, H. R. Mahmud, M. I. B. Saripan, A. R. B. Ramli, and B. Karasfi, "Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound," *Clinical Imaging*, vol. 37, 2013.
- [8] S. Shirmohammadi and A. Ferrero, "Camera as the instrument: the rising trend of vision based Measurement," *IEEE Instrumentation and Measurement Magazine*, vol. 17, pp. 41–47, 2014.
- [9] J. Suckling, "Mammographic image analysis society (mias) database v1. 21," 2015, <https://www.repository.cam.ac.uk/handle/1810/250394>.
- [10] J. Nagi, A. K. Sameem, F. Nagi, and S. K. Ahmed, "Automated breast profile segmentation for ROI detection using digital mammograms," in *Proceedings of the IEEE Conference on Biomedical Engineering and Sciences (IECBES) 2010*, Kuala Lumpur, Malaysia, November 2010.
- [11] R. Adams and L. Bischof, "Seeded region growing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, pp. 641–647, 1994.
- [12] S. Hore, S. Chakraborty, S. Chatterjee et al., "An integrated interactive technique for image segmentation using stack based seeded region growing and thresholding," *International Journal of Electrical and Computer Engineering*, vol. 6, 2016.
- [13] N. Gedik, "A method for the classification of mammograms using a statistical-based feature extraction," *International Journal of Biomedical Engineering and Technology*, vol. 38, pp. 1–108, 2018, p.
- [14] L. K. k. Et al and B. N. Jagadesh, "A review on feature selection techniques in digital mammograms," *Turkish Journal of Computer and Mathematics Education*, vol. 12, pp. 3329–3338, 2021, p.
- [15] O. N. Oyelade and A. E. Ezugwu, "A novel wavelet decomposition and transformation convolutional neural network with data augmentation for breast cancer detection using digital mammogram," *Scientific Reports*, vol. 12, pp. 5913–5922, 2022, p.
- [16] P. K. G. Diderot and N. Vasudevan, "A hybrid approach to diagnosis mammogram breast cancer using an optimally pruned hybrid wavelet kernel-based extreme learning machine with dragonfly optimisation," *International Journal of Computer Aided Engineering and Technology*, vol. 14, pp. 408–425, 2021, p.
- [17] D. J. Kalita, V. P. Singh, and V. Kumar, "Detection of breast cancer through mammogram using wavelet-based LBP features and IWD feature selection technique," *SN Computer Science*, vol. 3, pp. 175–214, 2022, p.
- [18] B. Sridhar, "A quality representation of tumor in breast using hybrid model watershed transform and Markov random fields," in *Proceedings of the 2020 International Conference on Computer Communication and Informatics (ICCCI)*, January 2020.
- [19] G. Meenalochini and S. Ramkumar, "Survey of machine learning algorithms for breast cancer detection using mammogram images," *Materials Today Proceedings*, vol. 37, pp. 2738–2743, 2021, p.
- [20] S. Z. Ramadan, "Methods used in computer-aided diagnosis for breast cancer detection using mammograms: a review," *Journal of Healthcare Engineering*, vol. 2020, pp. p–21, Article ID 9162464, 2020.
- [21] M. Veta, J. P. W. Pluim, P. J. Van Diest, and M. A. Viergever, "Breast cancer histopathology image analysis: a review," *IEEE Transactions on Biomedical Engineering*, vol. 61, pp. 1400–1411, 2014.
- [22] M. S. R. M. Simões, "Automatic breast density classification on tomosynthesis images," *Dissent*, <https://run.unl.pt/handle/10362/115386?mode=full>, 2021.
- [23] M. Adel, M. Rasigni, S. Bourennane, and V. Juhan, "Statistical segmentation of regions of interest on a mammographic image," *EURASIP Journal on Applied Signal Processing*, vol. 2007, Article ID 49482, 8 pages, 2007.
- [24] A. Elmoufidi, K. El Fahssi, S. Jai-Andaloussi, N. Madrane, and A. Sekkaki, "Detection of regions of interest in mammograms by using local binary pattern, dynamic K-means algorithm and gray level co-occurrence matrix," in *Proceedings of the 2014 5th International Conference on Next Generation Networks and Services (NGNS'14)*, Casablanca, Morocco, May 2014.

- [25] A. Elmoufidi, K. El Fahssi, S. Jai-Andaloussi, and A. Sekkaki, "Detection of regions of Interest in mammograms by using local binary pattern and dynamic K -means algorithm," *International Journal of Image and Video Processing: Theory and Application*, vol. 1, 2014.
- [26] A. Elmoufidi, K. El, S. Jai-andaloussi et al., "Automatic diagnosing of suspicious lesions in digital mammograms," *International Journal of Advanced Computer Science and Applications*, vol. 7, 2016.
- [27] U. K. Veena and V. Jayakrishna, "CAD based system for automatic detection et classification of suspicious lesions in mammograms," *International Journal of Emerging Trends et Technology in Computer Science (IJETTCS)*, vol. 3, 2014.
- [28] N. M. Basheer and M. H. Mohammed, "Classification of breast masses in digital mammograms using support vector machines," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, 2013.
- [29] S. U. Khan, N. Islam, Z. Jan, K. Haseeb, S. I. A. Shah, and M Hanif, "A machine learning-based approach for the segmentation and classification of malignant cells in breast cytology images using gray level co-occurrence matrix (GLCM) and support vector machine (SVM)," *Neural Computing & Applications*, vol. 34, no. 11, pp. 8365–8372, 2021, p.
- [30] J. Malek, A. Sebri, S. Mabrouk, K. Toriki, and R Tourki, "Automated breast cancer diagnosis based on GVF-snake segmentation, wavelet features extraction and fuzzy classification," *Journal of Signal Processing Systems*, vol. 55, no. 1-3, pp. 49–66, 2009.
- [31] A. Khamparia, S. Bharati, P. Podder et al., "Diagnosis of breast cancer based on modern mammography using hybrid transfer learning," *Multidimensional Systems and Signal Processing*, vol. 32, pp. 747–765, 2021.
- [32] R. Rajendran, S. Balasubramaniam, V. Ravi, and S. Sennan, "Hybrid optimization algorithm based feature selection for mammogram images and detecting the breast mass using multilayer perceptron classifier," *Computational Intelligence*, vol. 38, 2022.
- [33] M. A. Al-Masni, M. A. Al-Antari, J. M. Park et al., "Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system," *Computer Methods and Programs in Biomedicine*, vol. 157, pp. 85–94, 2018.
- [34] S. Timp, C. Varela, and N Karssemeijer, "Computer-aided diagnosis with temporal analysis to improve radiologists' interpretation of mammographic mass lesions," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 3, pp. 803–808, 2010.
- [35] B. Halalli and A Makandar, "Computer aided diagnosis-medical image analysis techniques," *Breast imaging*, vol. 85, 2018.
- [36] D. A. Zebari, D. A. Ibrahim, D. Q. Zeebaree et al., "Systematic review of computing approaches for breast cancer detection based computer aided diagnosis using mammogram images," *Applied Artificial Intelligence*, vol. 35, pp. 2157–2203, 2021, p.
- [37] J. A. Jaleel, S. Salim, and S. Archana, "Textural features-based computer aided diagnostic system for mammogram mass classification," in *Proceedings of the 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICT)*, Kanyakumari, India, July 2014.
- [38] P. Kaur, G. Singh, and P. Kaur, "Intellectual detection and validation of automated mammogram breast cancer images by multi-class svm using deep learning classification," *Informatics in Medicine Unlocked*, vol. 16, Article ID 100239, 2019.
- [39] M. Y. Kamil and A.-L. A. Jassam, "Analysis of tissue abnormality in mammography images using gray level co-occurrence matrix method," in *Journal of Physics: Conference Series* vol. 1530, IOP Publishing, Article ID 012101, 2020.
- [40] N. Al-Azzam and I. Shatnawi, "Comparing supervised and semi-supervised machine learning models on diagnosing breast cancer," *Annals of Medicine and Surgery*, vol. 62, 2020.
- [41] M. W. Huang, C. W. Chen, W. C. Lin, S. W. Ke, and C. F. Tsai, "SVM and SVM ensembles in breast cancer prediction," *PLoS One*, vol. 12, no. 1, Article ID e0161501, 2017.
- [42] H. Asri, H. Mousannif, H. A. Moatassime, and T. Noel, "Using machine learning algorithms for breast cancer risk prediction and diagnosis," *Procedia Computer Science*, vol. 83, pp. 1064–1069, 2016.
- [43] R. Rawal, "Breast cancer prediction using machine learning," *Journal of Emerging Technologies and Innovative Research*, vol. 7, no. 5, 2020.
- [44] Y. Kourdifi and M. Bahaj, "Applying best machine learning algorithms for breast cancer prediction and classification," in *Proceedings of the International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS)*, Kenitra, Morocco, December 2018.
- [45] H. Cai, Q. Huang, W. Rong et al., "Breast microcalcification diagnosis using deep convolutional neural network from digital mammograms," *Computational and Mathematical Methods in Medicine*, vol. 2019, Article ID 2717454, 10 pages, 2019.
- [46] G. V. Ionescu, M. Fergie, M. Berks et al., "Prediction of reader estimates of mammographic density using convolutional neural networks," *Journal of Medical Imaging*, vol. 6, 2019.
- [47] D. A. Zebari, D. A. Ibrahim, D. Q. Zeebaree et al., "Breast cancer detection using mammogram images with improved multi-fractal dimension approach and feature fusion," *Applied Sciences*, vol. 11, no. 24, Article ID 12122, 2021.
- [48] R. C. Gonzalez, *Digital Image Processing*, Pearson Education, India, 2009.
- [49] W. K. Pratt, *Introduction to Digital Image Processing*, CRC Press, Boca Raton, FL, USA, 2013.
- [50] R. Beichel and M. Sonka, "Computer vision approaches to medical image analysis," *Lecture Notes in, Computer Science*, Springer, Berlin, Germany, 2006.
- [51] R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786–804, 1979.
- [52] R. F. Walker, P. Jackway, and I. D. Longstaff, "Improving Co-occurrence matrix feature discrimination," in *Proceedings of the 3rd Conference on Digital Image Computing: Techniques and Applications (DICTA'95)*, Brisbane, Australia, 1995.
- [53] A. Kumar, S. Mukherjee, and A. K Luhach, "Deep learning with perspective modeling for early detection of malignancy in mammograms," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 22, pp. 627–643, 2019.
- [54] C. T. Leondes, *Medical Imaging Systems Technology Analysis and Computational Methods*, pp. 63–85, World Scientific, Singapore, 2005.
- [55] A. Silik, M. Noori, W. Altabey, R. Ghiasi, and Z Wu, "Comparative analysis of wavelet transform for time-frequency analysis and transient localization in structural health monitoring," *Structural Durability & Health Monitoring*, vol. 15, pp. 1–22, 2021.

- [56] S. P. Singh and S. Urooj, "Wavelets: biomedical applications," *International Journal of Biomedical Engineering and Technology*, vol. 19, pp. 1–25, 2015.
- [57] H. D. Cheng, X. J. Shi, R. Min, L. M. Hu, X. P. Cai, and H. N. Du, "Approaches for automated detection and classification of masses in mammograms," *Pattern Recognition Society*, vol. 39, pp. 646–668, 2006.
- [58] B. K. Elfarra and I. S. Abuhaiba, "Mammogram computer aided diagnosis," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 5, 2012.
- [59] A. Al-Hashedi, B. Al-Fuhaidi, A. M. Mohsen et al., "Ensemble classifiers for arabic sentiment analysis of social network (twitter data) towards covid-19-related conspiracy theories," *Applied Computational Intelligence and Soft Computing*, vol. 2022, Article ID 6614730, 10 pages, 2022.
- [60] K. Ganesan, U. R. Acharya, C. K. Chua, C. M. Lim, and K. T. Abraham, "One-class classification of mammograms using trace transform functionals," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 2, pp. 304–311, 2014.
- [61] A. Santos, E. Figueiredo, M. Silva, C. Sales, and J. Costa, "Machine learning algorithms for damage detection: kernel-based approaches," *Journal of Sound and Vibration*, vol. 363, pp. 584–599, 2016.
- [62] M. Shen, Lanxin Lin, Jialiang Chen, and C. Chang, "A prediction approach for multichannel EEG signals modeling using local wavelet SVM," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 5, pp. 1485–1492, 2010.
- [63] H. Wang, B. Zheng, S. W. Yoon, and H. S. Ko, "A support vector machine-based ensemble algorithm for breast cancer diagnosis," *European Journal of Operational Research*, vol. 267, no. 2, pp. 687–699, 2018.
- [64] L. Wei, Y. Yang, R. M. Nishikawa, and Y. Jiang, "A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications," *Medical Imaging*, vol. 24, 2005.