Applied Medical Informatics
Original Research Vol. 45, No. 2/2023, pp: 36-48

Automatic Application Watershed in Early Detection and Classification Masses in Mammography Image using Machine Learning Methods

Pascal VAGSSA[1*], Olivier VIDEME[2], Martin Luther PASCAL[1,3], Guidedi KALADZAVI[1], KOLYANG[1]

[1] LaRI Lab, University of Maroua, P.O. Box 814 Maroua, Cameroon
[2] The National Advanced School of Engineering of Yaoundé, P.O.Box 8390, Cameroon
E-mail(*): pvagssa@gmail.com;
[3] Department of Applied Mathematics and Computer Science, School of Geology and Mining Engineering, University of Ngaoundr, P.O Box 115, Meiganga, Cameroon
E-mail(*): pvagssa@gmail.com;

* Author to whom correspondence should be addressed; Tel: + (237) 677 302 439

Received: April 9, 2023 /Accepted: June 24, 2023/ Published online: July 1, 2023

Abstract

Radiologists use mammogram images for the diagnosis of breast cancer. However, interpreting these images remains challenging depending on the type of breast, especially on dense breasts. Dense breasts may contain abnormal structures similar to normal breast tissue and could lead to a high rate of false positives and false positives negatives. We present an efficient computer-aided diagnostic system for detecting and classifying breast masses. After removing noise and artifacts from the images using 2D median filtering, mathematical morphology and pectoral muscle removal by Hough's algorithm, the resulting image is used for breast mass segmentation using the watershed algorithm. After the segmentation, the system extracts several data by the wavelet transform and the co-occurrence matrix (GLCM) to finally lead to a classification as malign or benign mass via the Support Vector Machine (SVM) classifier. This method was applied to 48 Medio-Lateral Oblique (MLO) images from the image base (mini-MIAS). The algorithm showed a 87.5% classification rate, 92.59% sensitivity, and a specificity of 93.94%.

Keywords: Breast cancer; Computer-Aided Diagnosis; Segmentation; Classification; Support Vector Machine (SVM)

Introduction

Breast cancer is the leading cause of death among women worldwide. Over 2.2 million cases of breast cancer were recorded in 2020, making it the most common cancer [1]. In breast cancer imaging is the key to quality management and provides increasing data [2]. This approach highlights features and /or indicators that can assist in classification. The arrival of digital images is a real asset in many scientific fields. Indeed, the technological progress concerning X-ray tubes and films significantly improves the detection of lesions in the breast. Mammography remains the reference technique for breast exploration, it is effective in early detection of breast cancer and monitoring. Mammography allows the detection of abnormalities such as opacities and calcifications, which may indicate malign or benign lesions [3]. However, the quantity of images the radiologist has to interpret in a limited time is important and constitutes a difficulty for the interpretation. Computer-aided diagnosis can improve the accuracy of mammography interpretation and thus be used as a second reading to improve the analysis results of mammograms [4].

Several works are proposed in the context of mammography interpretation. Kom et al. [5] proposed a mass detection algorithm using a linear transformation filter algorithm to enhance the image, it subtracted the enhanced image from the original image to obtain a difference image. A local adaptive thresholding technique was developed to detect mass in the difference image. The algorithm accuracy for sensitivity rate was 95.91% and 93.87%, respectively, when the preprocessing step was or was not applied.

Xu et al. [6] proposed a segmentation algorithm for mamograms and investigated its performances on 363 lession using surface overlap metric (AOM), Hausdorff distance (HD), and mean minimum Euclidean distance (AMED). They respectively obtain the results The reported mean $\pm$ standard deviations were 0.72 ± 0.13 for AOM, 5.69 ± 2.85 mm for HD and 1.76 ± 1.04 mm for AMED.

Sapate et al. [7] propose a hybrid mechanism to automatically detect suspicious lesions using connected component labeling and an adaptive fuzzy region growth algorithm. A novel neighboring pixel selection algorithm reduces the computational complexity of the seeded region growth algorithm used to finalize lesion contours. The lesions are characterized using radiomic features and then classified as benign mass or malign tumour using k-Nearest Neig (k-NN) and Support Vector Machine (SVM) classifiers. The qualitative evaluation of the segmentation results by expert radiologists shows a sensitivity of 91.67% and a specificity of 58.33%. The overall features achieved a sensitivity of 84.44% and 85.56%, specificity of 91.11% and 91.67% with False positives per Image (FPsI) of 0.54 and 0.55 using k-NN and SVM classifiers on a local dataset respectively.

Kadhim et al. [8] presented a comprehensive algorithm for detecting abnormal masses by anatomical segmentation of the breast region of interest (ROI) on oblique medial-lateral mammograms (MLO). The authors proposed the marker-controlled watershed method. The proposed algorithms are fully autonomous, and are capable of isolating abnormal regions of breast tissue. The algorithm used in this work is marker dependent, which in turn depends upon the selected valueof threshold. In this work, the optimal value of threshold is selected interactively.

Husain et al. [9] proposed an improved watershed segmentation algorithm that uses Radial Basic unction (RBF) neural networks for the segmentation of image target objects. They use RBF neural networks to predict the boundaries of the ends of the segmentation clusters that are formed from the watersheds created in the image histogram topography. The initial RBF parameters, such as centres and widths, are automatically set to the histogram's peaks and minima.

Varsha and Gaikwad [10] present an approach based on the watershed algorithm, thus the main drawback of this algorithm is over-segmentation. To alleviate the over-segmentation problem, Varsha and Gaikwad [10] used the morphological marker-controlled method and gradient image for the watershed algorithm to minimize over-segmentation in the image analysis.

Soulami et al. [11] present a computer-aided diagnostic system for detecting and classifying ambiguous areas in dense breast mammograms. The noise and artifact were removed using 2D median filtering and labeling. The anomalies were isolated using electromagnetic type metaheuristic optimization (EML) algorithm. Finally, performs descriptor extraction using Zernike Moments to classify mammogram anomalies into normal or abnormal region via support vector machine (SVM).

Chattaraj et al. [12] introduced a new marker-controlled watershed algorithm for mammographic segmentation to highlight suspicious regions more distinctly. They used the morphological operations to obtain watersheds from a topographic demonstration of the input image. The proposed method has been applied and tested on the MIAS & BIRADS databases.

Sachin et al. [13] implemented a novel hybrid approach based on combining multi-level Otsu and watershed segmentation to extract the suspicious sections from the digital mammogram. Initially, multilevel thresholding using bat algorithm (BA) driven Otsu with two, three and four level thresholding is implemented to preprocess the digital mammogram. Then, a marker-controlled is implemented to extract the infected division from digital mammogram. The digital mammogram dataset with dense, medium, low and normal breast regions is analysed independently with the proposed approach. The proposed method proved effective in extracting breast malignancy from the considered digital mammogram database.

Embong and Anuar [14] used the marker-controlled watershed algorithm for the segmentation of abnormalities on mammograms. They applied filtering by principal component analysis (PCA). Before the segmentation by the watershed algorithm, the foreground of the image, which is the region of interest (region of anomaly), they identify the foreground of the image by the Fuzzy clustering algorithm C-Means (FCM).

Hmida et al. [15] present an approach for the automatic segmentation of breast mass. Their approach is mainly composed of three steps: initialization of the contour applied to a given region of interest; construction of fuzzy contours and estimation of fuzzy membership maps of different classes in the considered image; integration of these maps into the Chan-Vese model to obtain a fuzzy energy-based model which is used for the final mass delineation. This approach is evaluated using the mini-MIAS image base. The experimental results show that the proposed method achieves an average true positive rate of 91.12% with an accuracy of 88.08%.

Nayak et al. [16] propose the use of an image processing technique for segmentation and estimation of breast density by mammography images using a watershed algorithm. The proposed technique is tested with the publicly available MIAS dataset [17], and the resulting accuracy is comparable to state-of-the-art techniques available in the literature with improved computational efficiency.

Table 1 outlines some of the advantages and disadvantages of the state-of-the-art breast cancer segmentation and classification approaches. In light of this summary, watershed algorithm [6, 8, 9, 12-14, 16] has some advantage such as extraction of malignant areas, higher computational efficiency with good performance and its easy implementation. Although there are some drawbacks like over-segmentation problem, if the markers control the segmentation process this problem can be managed. Acceptable accuracy is the main advantage of SVM [7,11]. It is still necessary to improve the performance of SVMs compared to deep learning algorithms, because they have difficulty in identifying tumor cells. All these works deal with detection, some deal with classification of abnormality on mammograms and some deal with detection and classification. Many of these works have acceptable results, but some have opted to segment the pectoral muscle, which can also bias the interpretation of mammography because of its structure similar to breast masses and classification. Thus, we propose to address these two limitations with our method.

In this paper, we propose an algorithm to assist in the early detection of masses based on mammography images to segment and classify them.

Material and Method

The method of automatic segmentation/classification of mammographic masses proposed in this paper is described in Figure 1 and it is implemented in MATLAB.

It uses some images from the MIAS (Mammographic Image Analysis Society) image database comprising a set of 322 images in PGM format and corresponding to the left and right breast, freely available online for scientific purposes [17]. Each pixel is described as an 8-bit word. Each image is 1024 × 1024 in size. Our study used 48 images from the MIAS (Mammographic Image Analysis Society) image database. We used 32 images (16 images with benign tumors and 16 images with malign tumors) for the training set. We constituted a test base of 16 images (8 images with benign tumors and 8 images with malign tumors). We carried out three tests and each time, we changed the images:

- The first test consists of eight (8) with malign tumors and eight (8) images with benign tumors.
- The second test consists of six (6) images with malign tumors and ten (10) images with benign tumors.
- The third test is made up of eleven (11) images with malign tumors and five (5) images with benign tumors.

Table 1. Summaries of methods used for image detection and classification

Authors	Objective	Method used	Result
Kom et al. [5]	Detection and segmentation	Linear transformation filter algorithm & Local adaptive thresholding technique	Se: 95.91% and 93.87%, respectively, when the preprocessing step was or was not applied.
Xu et al. [6]	Segmentation of breast lesions	Watershed transformation	AOM, HD, and AMED of 0.72, 5.69 mm, and 1.76 mm, respectively.
Sapate et al. [7]	Segmentation and classification of breast lesions	Connected component labelling, Adaptive fuzzy region growth algorithm, Neighbouring pixel selection algorithm & radiomic features and using k-NN and SVM classifiers.	- Segmentation: Se=91.67, Sp=58.33 - Classification: Se=84.44% and 85.56%; Sp=91.11% and 91.67%. FPsI = 0.54 (k-NN) and 0.55 (SVM)
Kadhim et al. [8]	Detection and segmentation	Marker-controlled watershed method	Higher computational efficiency with good performance
Husain et al. [9]	Object segmentation in an image	Watershed algorithm & RBF neural networks	Effective tool to define suitable target regions markers, and it can contribute to overcome the conventional watershed segmentation problems.
Varsha and Gaikwad [10]	Mammography segmentation	Watershed algorithm &Watershed monitored by markers	Algorithm is superior to conventional algorithm in terms of over segmentation
Soulami et al. [11]	Detection and classification	EML algorithm, Zernike Moments & SVM Classification	Accuracy: 86.36; Sensitivity : 81.81 ; Specificity : 90.9
Chattaraj et al. [12]	Segmentation	Marker controlled watershed algorithm	The proposed method provides the actual shape and position of the masses in mammograms
Sachin et al. [13]	Segmentation and classification	Hybrid approach based on the combination of Otsu multi-threading and watershed	Very effective in extracting breast malignancy from the relevant DM database
Embong and Anuar [14]	Segmentation	Marker-controlled watershed algorithm, principal component analysis (PCA) filtering.	Three shapes of structuring elements, the disc, the diamond and the octagon, are tested and compared, they found that the diamond-shaped structuring element is a suitable shape for mammography image segmentation.
Hmida et al. [15]	Automatic segmentation	Fuzzy contours and estimation of fuzzy membership maps of different classes in the considered image, Chan-Vese model to obtain a fuzzy energy-based model.	The proposed method achieves an average true positive rate of 91.12% with an accuracy of 88.08%.
Nayak et al. [16]	Automatic segmentation of breast tumours	Watershed algorithm	Improved computational efficiency

Se = sensitivity; Sp = specificity; k-NN = ; SVM = support vector machine; RBF = Radial Basic Function

We aim to develop a Breast Cancer Computer Aid Diagnostic (CAD), based on the study of mammography images, and revolving around the following steps: (1) Preprocessing, (2) Segmentation, (3) Extraction, (4) Classification.

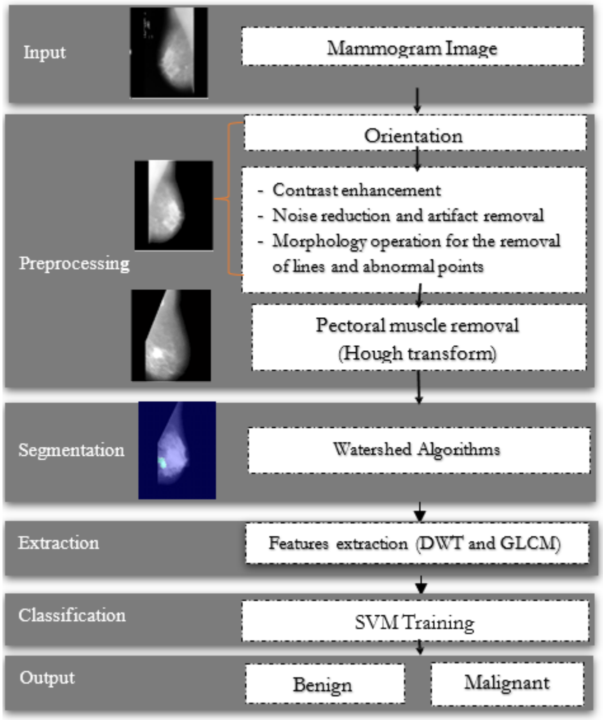

Figure 1. Diagram of proposed Computer Aid Diagnostic model

Preprocessing

The preprocessing is carried out step by step, starting with the orientation of the image, followed by the reduction of artefacts and various noises, the removal of the pectoral muscle closes the preprocessing [18].

Segmentation

Image segmentation is a technique that separates or divides the input image into different meaningful parts depending on the type of object. We usually segment regions by identifying common properties. Similarly, we need to detect contours using the differences between the different regions. The basic clue for identifying region separators is to use pixel intensity values. Therefore, the watershed technique constrained markers.

Watershed transformation can be qualified as a regional segmentation approach, proposed in 1977 by Digabel and Lantuejoul and later improved by Li et al. [19].

The basic idea behind the watershed segmentation method is to consider a greyscale image as a topographic relief. It is then a matter of calculating the watershed of this relief. The watersheds thus obtained correspond to the regions of the partition. It is then possible to define the watershed as the ridge forming the boundary between two watersheds. A catchment area being a geographical area, a drop of water following the line of greatest slope will arrive on a given minimum. A minimum is associated with a watershed.

Watershed Controlled-Markers

To use the watershed, transform in segmentation, one first preprocesses the greyscale image using the magnitude of the gradient, which is calculated using a linear filtering method [8]. Thus, for any image at coordinates (x, y), the magnitude of the gradient vector and the angle at which the maximum and rate of change of the intensity level occurs at the specified coordinates (x, y) can be calculated using the following equations:

$$\mathrm{I}(\mathrm{x},\mathrm{y}) = \sqrt{(\mathrm{I}_1^2(\mathrm{x},\mathrm{y}) + (\mathrm{I}_2^2(\mathrm{x},\mathrm{y})} \quad (1)$$

$$\alpha(x, y) = \tan^{-1}\left(\frac{I_1(x, y)}{(I_2(x, y)}\right) \tag{2}$$

where I_1 (x, y) and I_2 (x, y) are the gradients in the x and y directions, respectively. The magnitude of these gradients can therefore be calculated using the Sobel masks h_1 and h_2, which are defined by the following equations:

$$h_1 = [-1\ 0\ 1;\ -2\ 0\ 2;\ -1\ 0\ 1] \tag{3}$$

$$h_2 = [-1 - 2 - 1;\ 0\ 0\ 0;\ 1\ 2\ 1] \tag{4}$$

Segmentation by watershed generally leads to a problem of over-segmentation. To overcome this problem, an approach based on the marker concept is often introduced (marker-controlled watershed segmentation). The internal markers from the grey scale image are determined and the watershed transform of the distance transformed image of the internal markers are calculated. Regional minima are then imposed on the internal and external markers to modify the gradient of the image. Subsequently, the watershed transform is applied to the modified image gradient to produce the watershed lines. The resulting watershed lines are then superimposed on the original image to obtain the segmented image.

The algorithm steps for marker-controlled watershed segmentation are as follows:

(a) Compute a gradient of the image
(b) Open-Close by reconstruction
(c) Regional maxima (markers)
(d) Open-Close by Threshold reconstruction
(e) Superposition of markers on the original image
(f) Compute a Marker Controlled Watersheds

Feature Extraction

Radiologists use the images' shapes, grey levels and textural properties when analyzed mammograms to detect masses. Thus, in our support system, several data are extracted by the wavelet transform and the Gray Level Co-occurrence Matrix (GLCM).

The wavelet transform (2DWT) is a powerful feature extraction tool. It is used here to locate the necessary signal information to classify our mammogram images. Thus, we used the 2D discrete wavelet transform, which led to four sub-bands Low-Low (LL), High-Low (HL), Low-High(LH), High-High(HH), with a three-level decomposition of our image [20]. The wavelet approximations at the first and second levels are represented by LL1, LL2 respectively.

The statistical features were extracted using the grey level co-occurrence matrix (GLCM), also called grey level spatial dependence matrix which was introduced by Haralick [18]. In this method, textural features such as contrast (Con), correlation (Cor), energy (Ene), homogeneity (Hom), mean (Mean), standard deviation (Sd), entropy (Ent) and variance (Var), Root Mean Squared (RMS), Smoothness (Smmoth), Kurtosis (Kurt), Skewness (Skew), Inverse Difference Movement (IDM), were obtained from the LL and HL sub-bands.

Classification and Evaluation

Support Vector Machine (SVM) is a supervised learning technique that searches an optical hyperplane to separate two sample classes. SVM classifier is one of the best classifiers suggested by many researchers that can be chosen for breast mass classification from mammograms [11]. It is dimensionality independent, feature space independent and mapping the input data into a higher dimensional space. It is done by using kernel functions with the aim of obtaining a better data distribution.

In this work, a classification of mammograms using SVM is proposed and the features for analysis are extracted using wavelet transform decomposition and competition matrix (GLCM). The obtained features are given as input to the SVM classifier. Finally, the output of our classifier (SVM) gives us image with malignant or benign tumors.

In the performance evaluation strategy of the proposed classification method, the measures used are: (sensitivity, specificity, classification rate, error rate,and and confidence interval for accuracy) :

- Sensitivity is the probability that the test will be positive if the tumour is malignant.

$$\text{Se (\%)} = \frac{\text{TP}}{(\text{TP+FN})} \times 100 \tag{5}$$

- Specificity (Sp): Represents the probability that the test is negative if the tumour is benign.

$$\text{Sp (\%)} = \frac{\text{TN}}{(\text{TN+FP})} \times 100 \tag{6}$$

- Classification rate (Cr): Represents the probability that the tumour is well classified.

$$\text{Cr (\%)} = \frac{(\text{TN + TP})}{(\text{T N + FN + TP + FP})} \times 100 \tag{7}$$

- Error rate (Er): Represents the proportion of misclassifications.

$$\text{Er (\%)} = \frac{(\text{FN + F P})}{(\text{TN + FN + TP + FP})} \times 100 \tag{8}$$

- Confidence interval for accuracy (Ic):

$$\text{Ic(\%)} = \left[f - \frac{1}{\sqrt{n}}; f + \frac{1}{\sqrt{n}}\right] \tag{9}$$

$$f = \frac{\text{Number of Positive/Negative}}{\text{n}} \tag{10}$$

where TP, TN, FP and FN represent: TP (True Positive): a malignant mass classified as malignant; TN (True Negative): a benign mass classified as benign.; FP (False Positive): a benign mass classified as malignant; FN (False Negative): a malignant mass classified as benign; f(the observed frequency of True Positive/Negative in the sample); n: the sample size.

Results

Preprocessing

Mamogram images are enhanced, and then for images with pectoral muscle, the muscle is removed to avoid misinterpretation by the specialist (Figure 2).

The result in Figure 2 shows how the artifact (red circles) and pectoral muscle (red triangle) are removed from the image, leaving the region of interest.

Segmentation

The region of interest (masses) was segmented using the Watershed algorithm described in section marker-controlled watershed algorithm and applied to the preprocessed images followed by edge detection. Figures 3 shows the experimental results of this step.

The result of Figure 3. shows that the masses (red circle) are directly visible on the mammography, this figure is the result of the application of our method on the mammography after the preprocessing step (noise suppression and pectoral muscle.

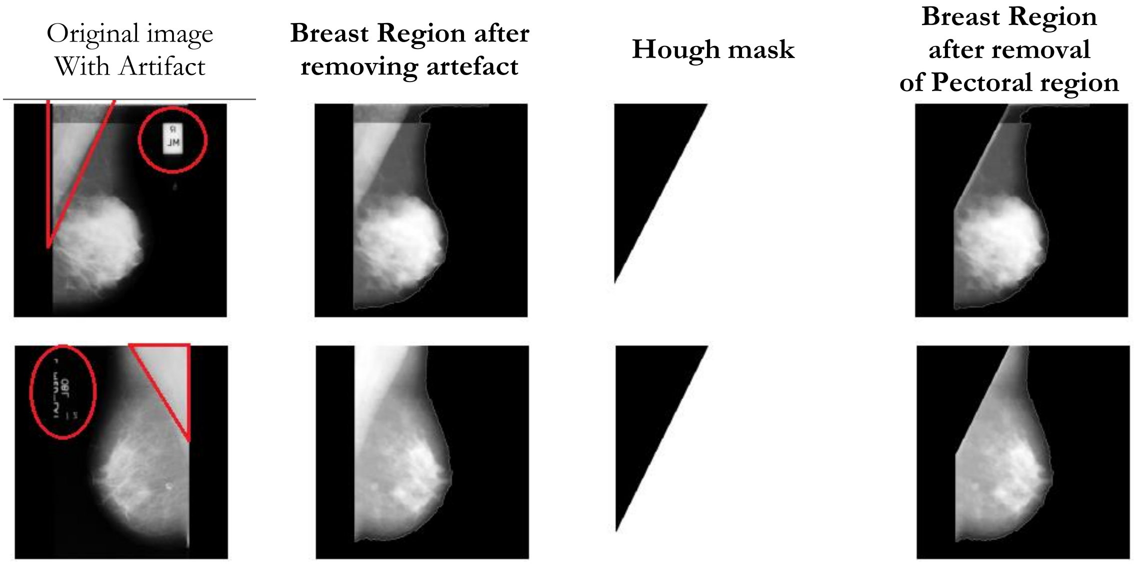

Figure 2. Preprocessing result

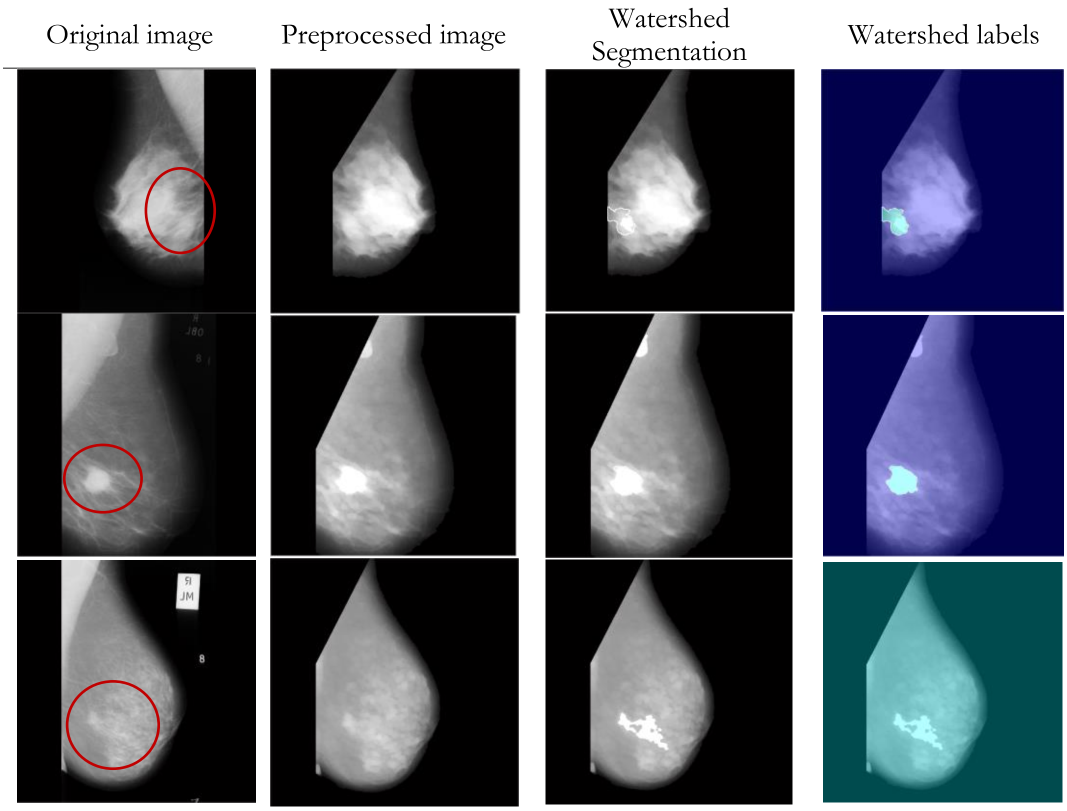

Figure 3. Segmentation result

Feature Extraction

The features obtained from the GLCM used to train our SVM classifier are presented in Table 2.1 and 2.2.

Table 2.1. The statistical features of trained images

Id Image	Con	Cor	Ene	Hom	Mean	Sd	Ent
mdb001	5.6309	0.5496	0.6132	0.8856	65.9593	295.1223	0.9461
mdb002	6.3564	0.5419	0.5668	0.8691	74.5224	312.6130	1.0815
mdb010	5.9910	0.5378	0.5972	0.8784	76.2918	315.7515	0.9891
mdb032	5.0608	0.5588	0.6547	0.9012	70.7477	324.2847	0.7927
mdb145	6.5117	0.5287	0.5796	0.8727	75.6258	305.6936	0.9436
mdb198	6.1913	0.5372	0.5891	0.8763	89.2425	357.1699	0.9669
mdb290	4.0562	0.5456	0.7230	0.9191	45.2532	253.5363	0.7035
mdb312	7.0426	0.5349	0.5318	0.8567	101.9983	372.6956	1.1160
mdb069	8.0019	0.5195	0.4843	0.8372	101.5916	346.6659	1.2132
mdb163	7.6794	0.5176	0.5072	0.8455	105.4750	389.3828	1.1542
mdb188	4.8857	0.5490	0.6665	0.9022	61.4517	293.4337	0.8249
mdb195	8.7447	0.4868	0.4609	0.8218	113.5071	381.0300	1.2949
mdb199	6.4999	0.5300	0.5741	0.8703	86.7700	345.4406	0.9906
mdb204	8.7343	0.4994	0.4492	0.8198	121.4881	388.4815	1.3262
mdb207	5.2745	0.5460	0.6390	0.8927	63.7590	291.4346	0.8918
mdb175	6.4948	0.5355	0.5735	0.8718	84.4577	341.8981	0.9634
mdb134	10.0020	0.4715	0.3941	0.7918	115.3349	347.1088	1.4738
mdb141	10.0175	0.4738	0.4053	0.7984	140.8236	432.4335	1.3500
mdb148	8.6026	0.4911	0.4768	0.8295	97.4527	343.2518	1.1633
mdb178	4.9711	0.5506	0.6638	0.9021	58.3626	283.4662	0.8024
mdb179	3.6261	0.5592	0.7342	0.9232	52.0847	298.0358	0.7507
mdb181	7.9070	0.4652	0.5309	0.8435	75.2578	309.6449	1.0954
mdb184	8.9179	0.4972	0.4411	0.8162	124.4714	398.1744	1.3383
mdb125	8.1026	0.5102	0.4962	0.8409	108.3106	384.0295	1.1084
mdb144	2.9318	0.4137	0.2849	0.7288	87.1625	80.3828	1.7528
mdb155	0.2683	0.4875	0.4324	0.8112	142.4325	438.6505	0.3447
mdb265	8.6042	0.5055	0.4568	0.8245	119.2017	96.4226	0.2765
mdb202	8.8886	0.5030	0.4448	0.8204	32.3301	18.6834	1.2760
mdb220	7.9070	0.4652	0.5309	0.8435	75.2578	309.6449	0.1629
mdb231	10.5486	0.4628	0.3793	0.7845	137.1431	400.8845	1.4641
mdb244	8.2275	0.5124	0.4692	0.8297	22.4545	401.9847	1.2775
mdb206	8.3348	0.5102	0.4755	0.8325	113.5833	384.7562	1.1978

Con = contrast; Cor = correlation; Ene = energy; Hom = homogeneity; Mean = arithmetic mean; Sd = standard deviation; Ent = entropy

Classification and Evaluation

Table 3 summarises the sensitivity, specificity, classification and error rates obtained by the classifier (SVM).

Table 3. Performance evaluation of the proposed method

Classifier (SVM)	Malign	Benign	TP	TN	FP	FN	Se (%)	Sp (%)	Cr (%)	Er (%)
Training	16	16	16	16	0	0	100	100	100	0
Test 1	8	8	7	8	0	1	87.5	100	93.75	6.25
Test 2	6	10	4	9	2	1	80	82	81.25	18.75
Test 3	11	5	9	5	0	2	82	100	87.5	12.5

Se = Sensitivity, Sp = Specificity, Cr = Classification rate, Er = Error rate, TP = True positive; TN = True Negative, FP = False Positive, FN = False Negative

The confidence intervals relating to images with malignant tumors and benign tumors are respectively [31.25; 81,25] and [25; 75].

Table 2.2. The statistical features of trained images

Id Image	Var	RMS	Smoot	Kurt	Skew	IDM
mdb001	3.9223e+04	99.4404	1.0000	27.0329	4.8699	1.9656e+05
mdb002	4.5787e+04	120.1643	1.0000	24.1179	4.5671	1.4665e+05
mdb010	3.9177e+04	109.8831	1.0000	18.9544	4.1314	2.0952e+05
mdb032	5.5741e+04	167.3754	1.0000	16.1512	3.7788	2.5231e+05
mdb145	4.9555e+04	128.8821	1.0000	4.0749	4.0749	1.6514e+05
mdb198	4.9555e+04	128.8821	1.0000	18.9660	4.0749	1.6514e+05
mdb290	3.5575e+04	79.6371	1.0000	37.0392	5.7908	1.6421e+05
mdb312	4.6602e+04	142.3702	1.0000	14.1543	3.5319	2.3239e+05
mdb069	4.3342e+04	140.1888	1.0000	13.7203	3.4050	2.3237e+05
mdb163	6.3662e+04	156.2056	1.0000	16.0293	3.7449	2.4695e+05
mdb188	4.5444e+04	105.4003	1.0000	27.1362	4.8911	1.8921e+05
mdb195	6.1427e+04	174.4240	1.0000	12.3796	3.2478	2.8389e+05
mdb199	3.8684e+04	119.2144	1.0000	18.0079	3.9827	1.8037e+05
mdb204	3.9352e+04	164.1936	1.0000	10.9459	3.0223	2.7256e+05
mdb207	3.8649e+04	97.0354	1.0000	24.2295	4.6698	1.6541e+05
mdb175	4.0923e+04	124.4414	1.0000	17.7965	3.9546	1.8505e+06
mdb134	4.7359e+04	180.8348	1.0000	10.2829	2.8353	2.0627e+05
mdb141	7.3080e+04	208.3348	1.0000	10.0334	2.9056	2.5651e+05
mdb148	5.4775e+04	62.1684	1.0000	14.6353	3.4902	2.2369e+05
mdb178	3.5885e+04	90.2346	1.0000	27.5408	4.9561	1.5269e+05
mdb179	5.0818e+04	93.3655	1.0000	34.8100	5.6659	1.4489e+05
mdb181	4.4158e+04	139.1915	1.0000	18.8175	3.9998	1.6604e+05
mdb184	4.0633e+04	159.1486	1.0000	10.8438	3.0499	2.8492e+05
mdb125	5.4915e+04	157.7586	1.0000	14.2673	3.5083	2.8146e+05
mdb144	5.1931e+04	235.1709	1.0000	6.9926	2.3510	3.5135e+05
mdb155	6.1819e+04	88.0247	1.0000	9.5242	2.8530	3.2366e+05
mdb265	4.7974e+04	07.6358	1.0000	19.3827	4.1330	.8965e+05
mdb202	3.9897e+04	167.2416	1.0000	10.5782	3.0002	.7345e+05
mdb220	5.3123e+04	67.9151	1.0000	3.0894	3.3057	2.6381e+05
mdb231	3.9757e+04	191.5580	1.0000	9.2325	2.7262	3.4767e+05
mdb244	3.6188e+04	132.0214	1.0000	18.0623	3.9962	2.3893e+05
mdb206	5.1623e+04	154.5613	1.0000	12.5715	3.2947	2.5164e+05

Var = variance, RMS = Root Mean Squared, Smmoth = Smoothness, Kurt = Kurtosis, Skew = Skewness, IDM = Inverse Difference Movement

Discussion

Our results showed that the proposed method appropriately distinguishes between malignant and benign masses which helps the radiologist to make a clear diagnosis and to take good decisions.

Mammography is mainly used for the diagnosis of breast abnormalities and/or breast cancer. An accurate diagnosis from mammograms can predict breast abnormalities. This requires a good interpretation of the mammograms to find the type of abnormality. Segmentation and classification algorithms such as Local Adaptive Thresholding Technique [5], Electromagnetic Metaheuristic Optimization (EML) and Mammography Anomaly Classification via Support Vector Machine (SVM) [11], Marker Controlled Watershed Algorithm for Mammography Anomaly Segmentation [13] are applied to mammography images for breast anomaly detection and classification. In this work, we presented a method to detect and classify breast masses.

The literature gives numerous studies on the diagnosis and classification of breast cancer based on machine learning in order to improve the accuracy of the systems in place (accuracy, sensitivity, specificity [11] or the false positive/negative rate [13]).

It should be noted that accuracy is very important in the diagnosis of breast cancer. Although we used here a sample image from the MIAS database, the technique can be applied to any data set.

From Table 3, we can see that the more true positives there are, the better the performance of our method in terms of sensitivity, and the more true negatives there are, the better the performance in terms of specificity. We can therefore say that the more malignant masses there are, the better our aid system will be able to detect them.

The segmentation and classification of masses by our method defined in Figure 1 on a set of 48 images (MIAS). According to the results obtained, our approach succeeds in detecting and classifying breast masses. To demonstrate the robustness of our approach, it was tested on mammograms with a training base of 32 images and a test base consisting of three tests of 16 images. Overall, for the evaluated mammograms, the accuracy reached a rate of 87.5%, a sensitivity of 92.59%, and a specificity of 93.94%, which means that the algorithm seems to be robust towards density types.

Note also that the 95% confidence intervals in relation to our results ([31.25; 81.25] and [25; 75] respectively the confidence interval relating to images with malignant and benign tumours) show in both cases that 95% of the results can be found in these intervals.

Table 4 summarizes some recent validated segmentation and classification work on mammography images. From this comparison, we can see once again that our method works well and gives satisfactory results for mass detection and classification.

The current study suffers from certain limitations, which should be corrected in future studies. First, we have the problem of image diversity. A variety of artificial intelligence methods may be the solution. Second, the image quality provided by the devices during image acquisition, which is one more step in processing. In future studies, more image samples should be considered. Third, we have not tested our method on another image database, this could be a hindrance. In future studies, assessments should be performed under multiple image bases.

Table 4. Comparison of the results obtained by our approach with those of the literature

Authors	**Classifier**	**Performance**		
		Acc (%)	**Se (%)**	**Sp**
Soulami et al. [9]	SVM	86.36	81.81	90.9
Sapate et al. [12]*	k-NN and SVM	n/a	84.44 (k-NN) vs. 85.56 (SVM)	91.11 (k-NN) vs. 91.67 (SVM)
Proposed method	SVM	87.5	92.59	93.94

*FPsI: 0.54 (k-NN) and 0.55 (SVM)

Conclusion

The proposed methodology provides accurate results of breast mass detection in mammograms with identification, localisation and classification of the type of mass. In the identification and classification into malignant and benign masses, an accuracy of 87.5% was achieved for the trained dataset as the statistical textural features were extracted from wavelet decomposition of LL and HL sub-bands and a sensitivity of 92.59% and a specificity of 93.94% was achieved for the dataset.

We plan to use different classifiers that can increase the accuracy by combining more efficient segmentation and feature extraction techniques with clinical cases using a large dataset covering different scenarios.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer. [Internet] 2023 [Cited 17 May 2023]. Available from: https: //gco.iarc.fr/today.
2. Boukhobza M, Mimi M. Détection Automatique de la présence d'anomalie sur une mammographie par réseau de neurones artificiels. Courier du Savoir. 2012;13:103-108.
3. Li H, Zhuang S, Li D, Zhao J, Ma Y. Benign and malignant classification of mammogram images based on deep learning. Biomedical Signal Processing and Control. 2019;51:347-354. doi:10.1016/j.bspc.2019.02.017
4. Dheeba J, Albert Singh N, Tamil Selvi S. Computer-aided detection of breast cancer on mammograms: a swarm intelligence optimized wavelet neural network approach. J Biomed Inform. 2014;49:45-52. doi: 10.1016/j.jbi.2014.01.010
5. Kom G, Tiedeu A, Kom M. Automated detection of masses in mammograms by local adaptive thresholding. Comput Biol Med. 2007;37(1):37-48. doi: 10.1016/j.compbiomed.2005.12.004
6. Xu S, Liu H, Song E. Marker-controlled watershed for lesion segmentation in mammograms. J Digit Imaging. 2011;24(5):754-63. doi: 10.1007/s10278-011-9365-2
7. Sapate SG, Mahajan A, Talbar SN, Sable N, Desai S, Thakur M. Radiomics based detection and characterization of suspicious lesions on full field digital mammograms. Comput Methods Programs Biomed. 2018;163:1-20. doi: 10.1016/j.cmpb.2018.05.017
8. Kadhim DA. Development algorithm-computer program of digital mammograms Segmentation for detection of masses breast using Marker-Control Watershed in MATLAB environment. Journal of Karbala University 2012;1:114-123.
9. Husain RA, Zayed AS, Ahmed WM , Elhaji HS. Image segmentation with improved watershed algorithm using radial bases function neural networks, 2015 16th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), Monastir. 2015; 121-126.doi: 10.1109/STA.2015.7505174.
10. Gaikwad VJ. Marker-Controlled Watershed Transform in Digital Mammogram Segmentation. International Journal for Research in Applied Science & Engineering 2015;3(3):18-21.
11. Soulami KB, Saidi MN, Tamtaoui A. A CAD system for the detection and classification of abnormalities in dense mammograms using electromagnetism-like optimization algorithm. International Conference on Advanced Technologies for Signal and Image Processing (ATSIP). 2017; pp. 1-8. doi: 10.1109/ATSIP.2017.8075533
12. Chattaraj A, Das A, Bhattacharya M. Mammographic image segmentation by marker-controlled watershed algorithm. 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2017, pp. 1000-1003, doi: 10.1109/BIBM.2017.8217793
13. Raj SPS, Madhava Raja NS, Madhumitha MR, Rajinikanth V. Examination of Digital Mammogram Using Otsu's Function and Watershed Segmentation. 2018 Fourth International Conference on Biosignals, Images and Instrumentation (ICBSII). 2018, pp. 206-212, doi: 10.1109/ICBSII.2018.8524794.
14. Rohana E, Anuar SR. Structuring Elements in the Watershed Algorithm for the Segmentation of Mammography Images. TENCON 2018 - 2018 IEEE Region 10 Conference, Jeju, Korea (South), 2018, pp. 2144-2147, doi: 10.1109/TENCON.2018.8650121.
15. Hmida M, Hamrouni K, Solaiman B, Boussetta S. Mammographic mass segmentation using fuzzy contours. Computer Methods and Programs in Biomedicine. 2018;164:131-142. Doi: 10.1016/j.cmpb.2018.07.005
16. Nayak T, Bhat N, Bhat V, Shetty S, Javed M, Nagabhushan P. Automatic Segmentation and Breast Density Estimation for Cancer Detection Using an Efficient Watershed Algorithm. In: Lecture Notes in Networks and Systems, Springer, Singapore. 2019;43:347-358. doi:10.1007/978-981-13-2514-4_29
17. Suckling J. The Mammographic Image Analysis Society Digital Mammogram Database. Exerpta Medica International Congress. 1994;1069:375-378.
18. Vagssa P, Doudou NM, Jolivo T, Videme O, Kolyang DT. Pectoral muscle deletion on a mammogram to aid in the early diagnosis of breast cancer. International Journal of Engineering Science and Technology. 2020;12(3):57-65. doi: 10.4314/ijest.v12i3.6

19. Li H, Elmoataz A, Fadili J, Ruan S. An improved image segmentation approach based on level set and mathematical morphology. In: Lu H, Zhang T (Eds). Proceedings of the Third International Symposium on Multispectral Image Processing and Pattern Recognition. 2003;5286:851-854.
20. Zhang YD, Wu L. An Mr Brain Images Classifier via Principal Component Analysis and Kernel Support Vector Machine. Progress In Electromagnetics Research. 2012;130:369-388. doi: 10.2528/PIER12061410
21. Haralick RM, Shanmugam K, Dinstein I. Textural feartures for image classification. IEEE Transactions on Systems Man and Cybernetics 1973;SMC3(6):610-621