

Detection and Classification of Breast Cancer in Mammography Images Using Pattern Recognition Methods

Naser Safdarian^{1,*}, Mohammad Reza Hediyezhadeh²

¹ Young Researchers and Elite Club, Tabriz Branch, Islamic Azad University, Tabriz, Iran

² Department of Biomedical Engineering, Dezful Branch, Islamic Azad University, Dezful, Iran

*Corresponding author: Naser Safdarian, Young Researchers and Elite Club, Tabriz Branch, Islamic Azad University, Tabriz, Iran. Tel: +98-9368173934; E-mail: naser.Safdarian@yahoo.com

DOI: 10.30699/acadpub.mci.3.4.13

Submitted: 4 August 2019

Revised: 29 August 2019

Accepted: 16 September 2019

e-Published: 1 October 2019

Keywords:

Classification

Breast Neoplasms

Electronic Data Processing

Image Processing, Computer-Assisted

Pattern Recognition, Automated

Introduction: In this paper, a method is presented to classify the breast cancer masses according to new geometric features.

Methods: After obtaining digital breast mammogram images from the digital database for screening mammography (DDSM), image preprocessing was performed. Then, by using image processing methods, an algorithm was developed for automatic extracting of masses from other normal parts of the breast image. In this study, 19 final different features of each image were extracted to generate the feature vector for classifier input. The proposed method not only determined the boundary of masses but also classified the type of masses such as benign and malignant ones. The neural network classification methods such as the radial basis function (RBF), probabilistic neural network (PNN), and multi-layer perceptron (MLP) as well as the Takagi-Sugeno-Kang (TSK) fuzzy classification, the binary statistic classifier, and the k-nearest neighbors (KNN) clustering algorithm were used for the final decision of mass class.

Results: The best results of the proposed method for accuracy, sensitivity, and specificity metrics were obtained $97\% \pm 4.36$, $100\% \pm 0$ and $96\% \pm 5.81$, respectively for support vector machine (SVM) classifier.

Conclusions: By comparing the results of the proposed method with the results of the other previous methods, the efficiency of the proposed algorithm was reported.

© 2019. Multidisciplinary Cancer Investigation

INTRODUCTION

Breast cancer is the most common cancer among women and is one of the major causes of cancer death in women aged 20 to 59 years. According to the Ministry of Health and Medical Education, it has become the most common disease in recent years in Iran [1]. Today, 88% of women diagnosed with breast cancer have a life expectancy of 10 years. In the United States, it has been reported that about 12% of women were identified during their lifetime, and were referred to as the second cause of women's death [2]. Early diagnosis of the disease is important because in the early stages

of the disease, cancer masses are restricted to the breast and the chance of surgical treatment in a less invasive manner is increased. The mortality rate is also decreased in the early stage. Also, the use of classifiers such as artificial neural networks in various fields of engineering sciences is increasing to analyze the time series and various issues of classification. Equipping medical science with intelligent tools for diagnosing and treating illness can greatly reduce physician errors and financial losses [1, 3]. Detection and diagnosis of the masses as irregular shapes accompanied by the changes

in size and occlusion of the breast tissue are very difficult [4]. Generally, signs and symptoms of breast cancer in mammographic images are classified into two categories including masses and calcium particles. Architectural distortions and focal asymmetry could be other signs of malignancy. These signs are present in mammographic images but their diagnosis is difficult and they need a high degree of radiologist accuracy. The masses are classified into benign and malignant categories. Benign masses have a very smooth and uniform margin of vision; while malignant masses have darker edges and highlights, and over time, these margins appear sharp and needle-shaped. The malignant masses are lobular or geometrically shaped like boots; while the benign masses have a circular or elliptical shape. The calcareous particles are very small calcium particles that appear as bright spots in mammographic images. The masses in the images appear with very low-intensity coefficients, and very small calcareous particles are usually confused with the noise particles in the image. The use of methods and techniques in the image processing field and the identification of patterns in the detection and classification of breast cancer from mammographic images reduces human errors and increases the speed of detection. Several studies have been conducted on the diagnosis and classification of breast cancer masses the most important of which are mentioned and reviewed in the following sections.

In 2013, Paramkusham Spandana et al., commenced the diagnosis of early-stage breast cancer; using novel imaging techniques by LabVIEW and MATLAB software [3]. In 2013, Pawar Punam et al., provided a recursive neural network model that provided an observer approach to training an artificial neural network for the diagnosis of breast cancer [1]. In 2013, Wang et al., tried to investigate an artificial neural network to predict the five-year survival of patients with breast cancer. Patients with breast cancer who were diagnosed and treated under standard treatment at a hospital in Taiwan between 2000 and 2003, were collected for artificial neural network testing [5].

In 2011, Abbosh Younis et al., presented a method for the diagnosis and locating the masses in the chest; using a neural network in sizes up to one-millimeter of radius; using four probes around the chest [6]. In 2013, Vadivel et al., presented the

classification of images with breast cancer masses; using the breast imaging reporting and data system (BI-RADS) database. They extracted geometric features from the cancerous masses and divided them into four different groups in terms of geometric shapes [7]. In 2016, Beheshti et al., used fractal methods to detect abnormalities in mammographic images. They used 168 carefully selected images by a radiologist as well as masses confirmed by biopsy. The images contained asymmetric lesions, architectural distortions, normal tissue, and mass tissue (include benign and malignant masses) [8].

In 2015, Anuj Kumar et al., provided a simple and easy way to detect masses in mammographic images. The proposed method was performed; using simple image processing techniques such as mean and threshold [9]. In 2010, Al-Shamlan Hala et al., presented a new method based on the values of the extracted features of mammographic images [10]. In 2015, Setiawan et al., studied the features of the rules as descriptions for categorizing mammographic images [11]. In 2018, Dhahbi et al., presented a novel method for reducing the false positives in the diagnostic system of mammogram images; using morphological analysis of tissue [12]. In that paper, the researchers used Hilbert's image representation, Kolmogorov-Smirnov distance, and maximum sub-region descriptors to improve the feature extraction and classification steps from the digital database for screening mammography (DDSM), and also used the feature selection.

In 2019, Chougrad presented a method for diagnosis of breast cancer on the Curated Breast Imaging Subset of DDSM (CBIS-DDSM), breast cancer digital repository (BCDR), INBreast, and MIAS (mammographic image analysis society) databases [13]. In 2018, Chakraborty et al., presented a method for automatic detection and diagnosis of mammographic masses and their categorization; using a multi-resolution analysis of tissue patterns [4]. In 2018, Rabidas et al., presented two new methods for feature extraction based on the neighborhood structural similarity (NSS) to determine the difference between benign and malignant masses in mammogram images [2]. In 2017, Shuyue Guan et al., developed a method for diagnosis of breast cancer based on transfer learning in CNN. Given that the CNN method requires a lot of labeled images, it initially extracts the features of the image that were obtained from

MIAS and DDSM databases and then categorizes them [14]. In 2018, Chatzistergos et al., presented a new method based on the combination of local binary pattern operators and radial lengths to classify the breast cancer masses. In that study, the researchers used MIAS and DDSM databases to obtain the results [15].

In this paper, several methods such as neural networks and fuzzy algorithms were developed for the diagnosis of breast cancer and predict the status of patients to reduce the diagnostic errors of experts and to make the diagnostic process cheaper. The availability and affordability of the methods presented in this paper suggest their application for early diagnosis of breast cancer. In fact, we analyzed and investigated mammographic images and determine the exact location of the mass with high accuracy; using image processing techniques. After the masses were extracted from mammographic images, several features were extracted from each mass. Finally, the extracted features were applied as the input of the classifiers and the results of the classifier were reported.

METHODS

In this section, we explain the different steps that were used in this paper for the detection and classification of breast cancer. In summary, the stages of image processing techniques include preprocessing, edge detection methods to detect the boundary of the cancerous masses, the separation of breast cancer masses from other healthy parts of the breast tissue, screening areas suspected to be a mass, extraction of features from the masses, construct the vector of features to apply to the classifier, and apply the vector features to the classifier. It should be noted that the type of the present study is applied research.

Introduce of Database

In this paper, the DDSM database was used to obtain breast cancer in mammogram images. This database includes normal texture, benign masses, and malignant masses [16]. The DDSM database produced by an organization from UK research groups is interested in researching mammographic images. This database is a producer of digital mammography images with 327 freely available images for download. The images in this database are 1024×1024 pixels. It should be noted that due

to the fact that cancerous masses are of breast thickness, often the main cause of the lesion and disease (which is about 3 to 30 mm in size) is extracted by the image processing techniques to detect the cancerous mass exactly.

Images Pre-Processing and Masses Detection

The pre-processing operation of the initial images taken from the DDSM database was performed to implement the algorithm for mammogram image processing [16]. The median filter is a nonlinear filter used to remove impulse noises. The initial image is often accompanied by noises or unwanted signals which sometimes causes the image to be erased or weakened. So removing these noise is essential. Median filtering scans all pixels of the image and uses local image processing methods by performing a masking action and adjusting the intensity of the brightness of adjacent pixels. This filter arranges all intensities of the ascending target pixel neighbors, selects the middle of the ordered numbers, and replaces the central pixel [17]. The output image of this filter is a smooth image with very little noise. A threshold method has been used to determine the boundaries of cancer masses in DDSM database images. The image threshold method is as follows. First, we divided the histogram in T_0 in the middle part. Then, for the brightness intensity ($T > T_0$ and $T < T_0$), we performed the mean and then we saw that the meanings were greater than the others. The proposed algorithm works in the same way



Figure 1: Extracting a Mass From a Sample Mammogram Image With Additional Areas

as in the first step to read the image from the input and store in the array $f(i,j)$. In the second step, we obtained the average brightness that was greater than the threshold (A) and smaller than the selected threshold (B). How to apply this algorithm is in equation 1.

Equation 1:
$$\begin{aligned} & \text{If: } f(i,j) \geq T_0. \text{ then } f(i,j) = 255. \\ & \text{elsef}(i,j) = 0 \end{aligned}$$

As a result, group A had a maximum brightness intensity, and group B had a minimum brightness intensity. Figure 1 is the result of using the threshold method in the mammography image [11]. But as shown in this figure, some additional areas along with the mass are also extracted. In the next step, the proposed algorithm was explained to remove additional detected areas.

Since the extracted mass is focused on the inside, the intensity of the pixels around the edge should be expanded to the center of the detected masses. Using the image morphological reconstruction method, we create a circle with an ascending radius that finds the intensity of its brightness by scrolling the image. After the repetition in the search, the radius was increased as soon as the target pixel was found. All pixels between the two areas had a new intensity equal to the intensity of the edge pixels. The results obtained by applying this algorithm are shown in Figure 2.

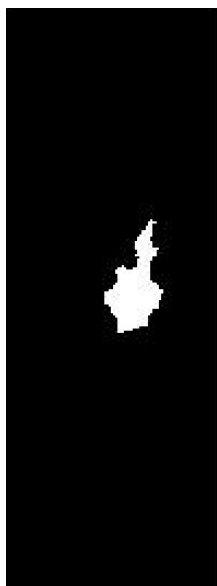


Figure 2: Exaction of a Cancerous Mass From a Sample Mammogram Image

Feature Extraction From Detected Masses

In this section, we examined the features extraction from detected cancer masses from mammographic images in MATLAB software. In this way, we analyzed binary images that represented the masses extracted by any dimension. In this paper, features such as area, bounding box, centroid, convex area, convex hull, convex image, eccentricity, Equiv diameter, Euler number, extent, extrema, filled image, filled area, image, major axis length, minor axis length, orientation, perimeter, pixel index list, pixel list, solidity, weighted centroid, sub-array index, thinness ratio, elongation (EN), circularity1, circularity2, compactness, distortion and disorientation (DP), and shape index (SI) were extracted from each image containing the extracted mass. Description of some of the extracted features are as follows:

The area feature of the mass is the actual number of pixels in the extracted area that determines the area of the cancer mass. This feature is a scalar number. The bounding box feature is the smallest rectangle into the extracted region which is a four-digit number. The first two numbers represent the range of the beginning of the rectangle (from the upper left corner of the area) which is x and the y column, respectively. The second two numbers represent the length of the region in the direction of x and y, respectively. The centroid feature is the vector that represents the center of the extracted mass. The first element of this vector is the characteristic of the coordinates x and the second element of the y coordinate. Figure 3 shows the features of the centroid and bounding box. The extracted region is shown in white pixels, the green box represents the bounding box feature, and the red dot represents the centroid feature.

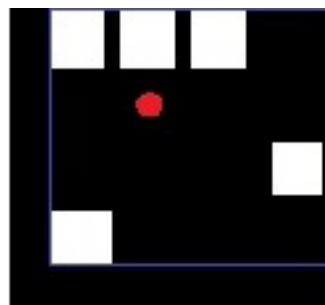


Figure 3: The Centroid and Bounding Box Features

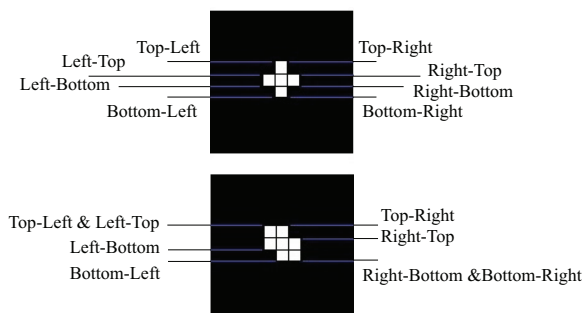


Figure 4: The End of Two Different Areas

The convex matrix body feature represents the smallest convex polygon that can be placed in the specified environment. Each row in the matrix containing the coordinates x and y is a polygon vertex. This feature is only supported for the two-dimensional (2D) input matrix. Convex image feature is a binary image (logical) that is defined as a convex body with all pixels of a full body. The image is as large as the area of the box. This feature is only supported by the 2D input matrix. A convex area feature is a number that represents the number of pixels with one value in the convex image. This feature is obtained by a 2D input matrix (a result is a number). An eccentricity feature is a number indicating the exit of the specified ellipse center which is a secondary region of the target area. This feature of the center exit is the ratio of the distance between the elliptical focus and its axis length. This feature has a value between zero and one (the values of zero and one are not used; if the number is zero, it means that the ellipse is exactly circular; whereas an ellipse with a center exit equal to one, is a tile line). This feature can be calculated in the form of a 2D label matrix as input. The Equiv diameter feature is a numerical value representing the diameter of a circle with the same region as the specified region. This feature is calculated as root by the expression $4 * \text{Area}/\pi$. This feature can be calculated in the form of a 2D label matrix as input. The Euler number feature is a numeric value that specifies the number of objects in the area minus the number of holes (or holes) in those objects. This feature is only supported by the 2D input matrix. The regionprops values use eight connectivity to calculate the measured value of the Euler number. The extent feature is a numerical property that specifies the ratio of pixels in the

region to pixels in the entire bounding box (which represents the smallest rectangle in the region). This calculation is obtained by dividing the area features into the bounding box region. This feature is only supported by the 2D input matrix.

The extreme feature is an 8×2 matrix that identifies the endpoints in the area under investigation. Each matrix row contains the coordinates x and y of a point. The vector format is top-left, top-right, right-top, right-bottom, bottom-right, bottom-left, left-bottom, and left-top. Figure 4 shows the end regions of two different regions. In the left-hand region, each endpoint is distinct. In the right-hand region, certain endpoints (for example, top-left and left-top) are the same.

The filled area feature is a scalar value that specifies the number of pixels contained in the filled image. The filled image feature is a binary image (logical) of the same size as the bounding box for the area in question. This feature is related to the bright pixels (on) of the region concerned with all the holes in it. Figure 5 shows the on and off pixels in this area.

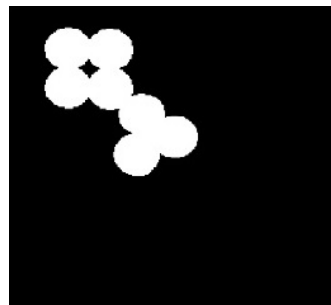


Figure 5: Displays the Pixels On and Off

The image feature is defined as a binary image (logic) that is the same size as the bounding box of the intended area. This feature is related to the bright pixels of the target area and all other pixels on it. The major axis length feature is defined as the length of the scalar value (in pixels) of the large elliptical axis which is defined as the normal secondary central region associated with the intended area. This feature is only supported by the 2D input matrix. The max intensity feature is a scalar value that represents the number of pixels that have the largest brightness in the area. The mean intensity feature is a scalar value that defines the average of all brightness values in the desired area. The min-intensity feature is a scalar value that represents the number of pixels that have the

smallest brightness in the area. The minor axis length feature is defined as a scalar value (in pixels) for the small elliptical axis that is defined as the normal secondary central region associated with the intended area. This feature is only supported by the 2D input matrix. The orientation feature is for a scalar angle (in degrees from -90° to 90°) defined between the x-axis and the large axis of the ellipse which is the same second moment of the region. This feature is only supported by the 2D input matrix. Figure 6 shows the axis and orientation of the corresponding ellipse. The left part of Figure 7 represents a region of the image and its ellipse. The right part shows the same ellipse with the graphic features as follows:

- Bold blue lines represent the axes;
- The red dots (which are defined in an ellipse in the right shape) indicate the focus;
- The orientation between the horizontal points and the big axis;

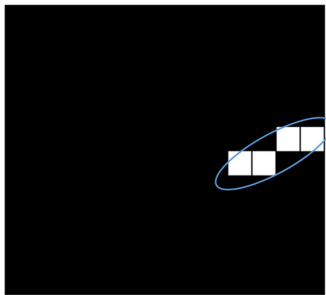


Figure 6: The Feature of the Direction That Is a Scalar Angle (in Degrees From -90° to 90°), Is Between the X-Axis and the Large Axis of the Ellipse

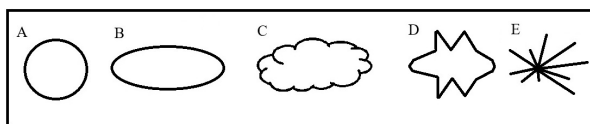


Figure 7: The Shape of Different Types of Breast Cancer [7]
A) Round mass; B) Oval mass; C) Lobulate mass; D) Irregular mass; E) Architectural distortion mass

The perimeter feature is a scalar value that represents the distance from the boundary of the target area. The regionprops toolbox calculates the spacing between a pair of adjacent pixels around the boundary of the target area to compute the perimeter feature. If the image contains segregated areas, this toolbox generates unexpected results. The pixel index list feature is a vector with a “p” element containing a linear index of pixels in the

target area. The pixel list feature is a P * Q matrix representing the position of the pixels in the target area. Each row of this matrix is [x y z ...] which specifies the coordinates of a pixel in the desired area. The pixel values feature is a P*1 matrix in which P is the number of pixels in the target area. Each element in this vector contains a value of one pixel in the region. The solidity feature is a clear scalar value that specifies the ratio of pixels in the convex sections of the body that are also in the area. This feature is calculated as the area/convex area relationship. This feature is only supported by the 2D input matrix. The subarray index feature is an array cell containing an index so that L(idx) extracts L elements within the object in the bounding box.

The weighted centroid feature is a P*Q vector in which the coordinates of the center of the area are obtained based on the position and amount of intensity. The first element of this feature is the horizontal coordinates (or x coordinates) of the center of gravity. The second element of this feature is the vertical coordinates (or y coordinates). All other elements of this feature are in other dimensions. The thinness ratio feature distinguishes circular masses from other types of masses. Of course, this ratio can be only used for circular masses. Equation 2 expresses this feature as follows:

$$\text{Equation 2: } TR = \frac{4\pi \times (\text{Mass Area})}{(\text{Mass Perimeter})^2}$$

The elongation (En) feature represents the difference between regular ellipsoidal masses with irregular elliptic masses. The elongation feature is the ratio of the minimum size to the maximum of a rectangle and is calculated by the equation 3.

$$\text{Equation 3: } En = \frac{\text{Mass Area}}{(2 \times \text{Max Radius of Mass})^2}$$

The circularity1 feature is defined in terms of the rotational dimension and the position of a mass which indicates how much a mass is similar to a circle. This factor is used to distinguish between circular and elliptic masses with irregularly deposited masses. The Circularity1 can be used for complete annular shapes and can be calculated from the equation 4 as follows.

$$\text{Equation 4: } C1 = \sqrt{\frac{\text{Mass Area}}{\pi \times (\text{Max Radius of Mass})^2}}$$

Circularity 2 feature indicates how much an extracted mass is similar to an elliptical mass (as reference). This factor is used to distinguish between elliptical or circular masses and irregular masses. The size of this feature is always equal to 1 for annular and squared squares. This feature is expressed by equation 5.

$$\text{Equation 5: } C2 = \sqrt{\frac{\text{Min Radius of Mass}}{\text{Max Radius of Mass}}}$$

The compactness feature is the degree of deviation of a complete cyclic mass in which this degree of deviation is independent of the linear variation of the mass. The hardness of a mass is obtained from the compression factor. The compaction feature is computed by equation 6.

$$\text{Equation 6: } CN = \frac{2 \times \sqrt{\text{Mass Area} \times \pi}}{\text{Mass Perimeter}}$$

The distortion and disorientation (DP) features are the characteristics of an irregular shape. The disorder of a mass is derived from the characteristics of the dispersion and diffusion of a mass. The dispersion and disorder features are determined by equation 7.

$$\text{Equation 7: } DP = \frac{\text{Max Radius of Mass}}{\text{Mass Area}}$$

The feature of the shape index (SI) is an indicator of the curvature of the mass surface. This feature gives us the edge profile of the mass and distinguishes it from the lean and dense powers of regular masses. This feature is defined by equation 8.

$$\text{Equation 8: } SI = \frac{\text{Mass Perimeter}}{2 \times \text{Max Radius of Mass}}$$

It should be noted that if the input image is a gray level image, the regionprops toolbox in the MATLAB software can calculate the max intensity, min intensity, mean intensity, pixel values, and weighted centroid features. But, since the image given to the input of this function is a binary image, the above features for this image are not calculated correctly. In this study, 19 features were selected and extracted from each image and used to generate the feature vector. These features consist of area, convex area, eccentricity, Equiv diameter, Euler number, extent, filled area, major axis length, minor axis length, orientation, perimeter, solidity, thinness ratio, elongation (EN),

circularity 1, circularity 2, compactness, distortion and disorientation (DP), and shape index (SI).

Classification

After extraction, the expressed features from the detected masses, all extracted features of each image were stored in a vector. By these vectors, we created the features matrix and this matrix was applied as a feature set for the input of the classification system. In the following step, we described the classification methods used in this paper. The neural network classifiers such as RBF, PNN, and MLP as well as the KNN clustering algorithm and SVM classifier (using RBF kernel) were used to classify the type of cancer masses in the breast tissue. Also, the TSK fuzzy and the Navie Bayes classifiers were used to determine the benign or malignant masses. In continue, we will describe some methods of classifiers used in this paper.

The probabilistic neural networks (PNN) classifier is a kind of RBF network. PNN neural networks were used in classification problems [18]. When an input vector is applied to the network, the first layer calculates the distance of the input vector from the inputs of the training data and thus provides a vector the elements of which will determine the distance between the input and input as training data. The second layer that uses the output of the first layer will produce the probability vector as the output of the network. The competitive transfer function (Compete) in the second layer, can be selected as the maximum likelihood of the vector probability. For this situation, the output is to produce 1, and for other probability the output is to produce zero [18]. The SVM was generally used for some problems with two categories. In this classifying method, two planes were put on the boundary of two data classes and the problem solved the maximum boundary between these two planes and between these two classes. In fact, machine learning is a model with supervising training that uses it for classification and regression data. The SVM is a binary linear classification. All of the classification methods used in this paper, classified type of the extracted masses, and the results of them are presented in the results section. It should be noted that the types of cancer masses in the breast tissue mammograms in the DDSM database were introduced as lobulated margins, oval, amorphous distribution, irregular

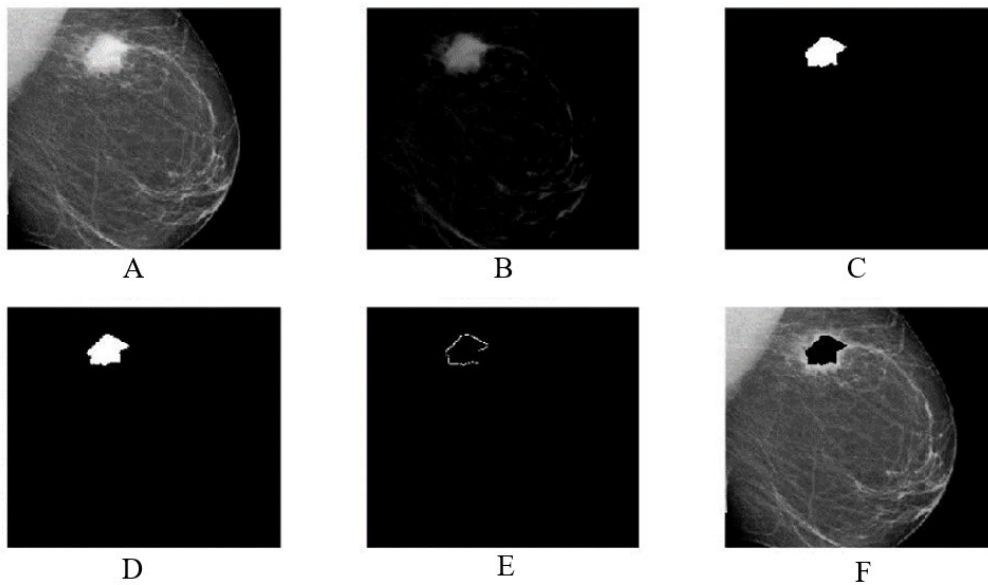


Figure 8: Results of Cancer Masses Extraction in a Mammography Image Taken From the DDSM Database

A) Initial image from DDSM database; B) Image pre-processing step; Noise reduction; Contrast enhancement; C) Extraction of cancerous mass; D) Delete points other than the mass of the image extracted; E) Extraction of the edge of the extracted mass; F) The cancerous mass is shown more clearly than the original image.

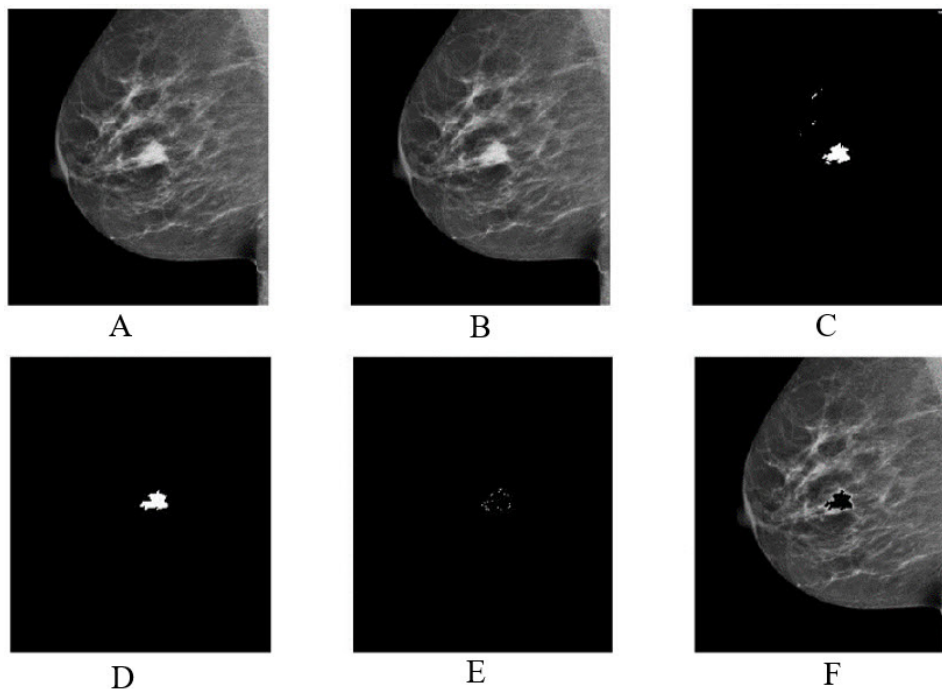


Figure 9: Results of Extraction of Cancer Masses in a Sample Mammography Image Taken From the DDSM Database

A) Initial image from DDSM database; B) Image pre-processing step; Noise reduction; Contrast enhancement; C) Extraction of cancerous mass; D) Delete points other than the mass of the image extracted; E) Extraction of the edge of the extracted mass; F) The cancerous mass is shown more clearly than the original image.

architectural, irregular margins, and pleomorphic distribution and round margins [16]. In the meantime, benign masses are shaped like circles or ovals, and the malignant masses are disordered and contain small, needle-shaped, and irregular appendages.

Figure 7 shows various shapes of breast cancer masses.

RESULTS

After applying the algorithm proposed in this paper

for the detection and classification of the breast cancer masses from mammogram images, results of the proposed algorithm, are reported. Figures 8 and 9 present the results of the proposed algorithm in this paper. The first images of the DDSM database are displayed in the first images of Figure 8 and Figure 9 [16]. The results of the cancer mass detection in each mammogram image are shown in these figures as well. These results contain the image preprocessing steps including noise deletion and recovery image contrast, extraction of cancer masses, removing additional regions that are not in the detected masses, edge extraction of the masses, and ultimately the final image in which the cancer masses have been shown more clearly than the original image.

After extraction of the cancer masses from all the images, all the features that were described in the previous section were extracted from all the detected masses and matrix features were applied to the input of the classification system. In this research, the feature matrix consisted of 19 columns and the number of columns in this matrix denotes the number of final extracted features from each detected mass from the images that were obtained from the database. To validate the proposed algorithm in this paper, the 4-Fold cross-validation method has been used. In this way, all existing data were divided into four sections and at each stage of the code, three parts of the

data were given as training data and one section as the validation data to run the classifiers. In this research, 85% of the data were considered as train data and 15% of the data were considered as test data. It should be noted that the runs of the code were carried out for a total of 100 times and finally the averaged and standard deviation of results were obtained and reported. In Table 1 and its subsequent equations (Equations 9, 10, and 11), the definition of sensitivity, specificity, and accuracy metrics are presented for classifying the mammogram images for diagnosis of the type of breast cancer mass.

$$\text{Equation 9: Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Equation 10: Specificity} = \frac{TN}{TN + FP}$$

$$\text{Equation 11: Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

The accuracy of the classification, as a result, is the number of correct detection algorithms in expressing the type of cancer mass divided into the total number of examined data. The accuracy of TSK fuzzy classifier, PNN neural network, the MLP neural network, the KNN clustering algorithm, the Bayes classifier, and SVM classifier for classifying the type of mass as benign or malignant were obtained $85.4\% \pm 0.9261$, $90.5\% \pm 3.0748$, $86\% \pm 3.513$, $83.33\% \pm 1.14$, $89\% \pm 3.35$, and

Table 1: Reference Labels and Labels Introduced in the Proposed Algorithm to Calculate the Sensitivity and Specificity ^a

	Algorithm Label		
		Benign	Cancer
Reference Label	Benign	Benign- Benign (TP)	Benign-Cancer (FN)
	Cancer	Cancer- Benign (FP)	Cancer-Cancer (TN)

^a Abbreviations: FN, False Negative; FP, False Positive; TN, True Negative; TP, True Positive

Table 2: Results of Accuracy, Sensitivity, and Specificity of the Proposed Method

Method	Accuracy	Sensitivity	Specificity
Bayes Classifier	$89\% \pm 3.35$	$95.556\% \pm 3.12$	$92.12\% \pm 2.98$
Probabilistic Neural Network (PNN)	$90.5\% \pm 3.07$	$100\% \pm 0$	$92.3\% \pm 1.15$
K-Nearest Neighborhood (KNN)	$83.33\% \pm 1.14$	$100\% \pm 0$	$81.82\% \pm 1.14$
Multi-Layer Perceptron (MLP)	$86\% \pm 3.51$	$99.3194\% \pm 2.50$	$89.03\% \pm 1.84$
Takagi-Sugeno-Kang (TSK) Fuzzy	$85.4\% \pm 0.93$	$82.6424\% \pm 0.71$	$81.28\% \pm 0.98$
Support Vector Machine (SVM)	$97\% \pm 4.36$	$100\% \pm 0$	$96\% \pm 5.81$

Researcher	Year	Method	Database	Accuracy	Sensitivity	Specificity
Punam S. Pawar [1]	2013	BPNN	Wisconsin breast cancer database	99%	—	—
Tan-Nai Wang [5]	2013	BPNN	database of breast cancer patients at a regional hospital in northern Taiwan	85.1%	94.83%	52.86%
A. Vadivel [7]	2013	fuzzy rules	DDSM	ILOR: 87.76%	—	—
				RO: 100%		
				RL: 95.45%		
S.M.A. Beheshti [8]	2016	detection asymmetric masses by SVM	MIAS	94.01%	91.54%	97.14%
		detection architectural distortion by SVM		95.29%	97.77%	93.75%
		detection CB and SM masses by SVM		95.30%	97.78%	95.84%
Arden Sagiterry Setiawan [11]	2015	law's texture energy measure (LAW's)	MIAS	Normal & Abnormal: 93.90%	Normal & Abnormal: 91%	Normal & Abnormal: 100%
		GLCM		Benign malignant: 83.30%	Benign malignant: 80%	Benign malignant: 88%
				Normal & Abnormal: 73.50%	Normal & Abnormal: 72%	Normal & Abnormal: 100%
				Benign malignant: 72.20%	Benign malignant: 75%	Benign malignant: 50%
Jayasree Chakraborty [4]	2018	multi-resolution analysis of oriented tissue patterns	DDSM	77.89%	—	—
Rinku Rabidas [2]	2018	NSS	mini-MIAS & DDSM	85.42%	—	—
Sevastianos E. Chatzistergos [15]	2018	combination of local binary pattern operator and radial lengths	MIAS	82.54%	50%	95.56%
Best Results of Proposed Method	2019	SVM classifier	DDSM	82.5%	80%	85%
			DDSM	97±4.36%	100±0%	96±5.81%

^a Abbreviations: BPNN, back propagation neural network; CB, circumscribed benign; DDSM, digital database for screening mammography; GLCM, gray level co-occurrence matrix; I, irregular; L, lobular; MIAS, mammographic image analysis society; NSS, neighborhood structural similarity; O, oval; R, round; SM, spiculated malignant; SVM, support vector machine

97%±4.36; respectively. The results of the accuracy, sensitivity, and specificity of the proposed method are presented in Table 2.

Finally, after applying the proposed algorithm, by comparing the results of the proposed method with other available methods, the accuracy, sensitivity, and specificity of the proposed algorithm were reported. Table 3 shows the comparison of the results of the proposed model with the results of previous studies.

DISCUSSION

As the early diagnosis of breast cancer can improve its therapeutic outcomes and thus increase the survival time of patients with breast cancer, development, and presentation of methods for diagnosing breast cancer are essential [14]. Therefore, in this article, a method was presented for the diagnosis of this cancer. In this paper, new features based on the geometric shape of the breast cancer masses were presented for the classification of these masses. In this research, after obtaining digital mammogram images from the DDSM database, all images were pre-processed (including noise deletion, imaging filters, and image cropping). Then, by applying the threshold technique on each image, we extracted the masses and their boundaries as well as the features of the extracted masses.

In this paper, after consulting with a radiologist expert, some of the above-mentioned features were removed from the research. Accordingly, the filled area features were extracted from the filled image feature, so this feature was removed. Also, the bounding box and image features that were related to the smallest rectangular in the extracted region were removed from this study. In addition, the subarray index features that extract the elements in the bounding box feature were removed according to the radiologist idea. Additionally, the convex area feature which was obtained from the convex image feature, was also removed. Finally, 19 features were introduced for extracting cancer masses from the mammogram images. In this study, 19 features consist of area, convex area, eccentricity, Equiv diameter, Euler number, extent, filled area, major axis length, minor axis length, orientation, perimeter, solidity, thinness ratio, elongation (EN), circularity1, circularity2, compactness, distortion and disorientation (DP), and shape index (SI) was extracted. Then each mass was introduced by its own feature vector; containing all the features extracted from the mammography

image's masses. The extracted features of digital mammography images stored in the matrix were used as input classifiers. In this paper, we used neural network classifiers such as RBF, PNN, and MLP types, as well as TSK fuzzy classifiers, the Bayes classifier, the KNN clustering algorithm, and SVM classifier for determining the final mass classification. Finally, after applying the proposed algorithm, by comparing the results of the proposed method with other available methods, the accuracy, sensitivity, and specificity of the proposed algorithm were reported. Table 3 shows the comparison of the results of the proposed model with the results of previous studies. According to Table 3, the proposed method in this paper has a better performance for the accuracy, sensitivity, and specificity parameter (the highest metrics is related to the SVM classifier) as compared to other similar previous studies. As shown in Table 3, the sensitivity results of the method presented in this paper were 100% which is higher than all previous studies. Therefore, this result could be the strength of the algorithm presented in this paper. Also, the accuracy result of the proposed method in this paper was 97±4.36% which was better than all previous researches except for the method presented in [1] in which this value was found to be 99%. It may be concluded that the results of this investigation cannot be compared with that paper as the database in [1] was different from the database that was used in this article. So this result has been compared with other previous papers and its superiority has been proved. The comparison of the specificity parameter with previous studies that used exactly the same database [2, 4, 7, 15] revealed that the outcomes of the present study are better than the results of previous researches. Regarding the above-mentioned discussion, the superiority of the presented algorithm in this paper is clearly established.

The innovations contained in this research include the high number of simple extracted features from the masses of mammogram images as well as the use of several methods for classifying mass types in order to determine the masses more precisely and with high precision.

ACKNOWLEDGEMENTS

The authors of this paper are grateful to the Young Researchers and Elite Club of Islamic Azad University of Tabriz.

CONFLICT OF INTEREST

There is no financial interests or potential conflicts of interest.

ETHICS APPROVAL

This article does not contain any studies with human participants and/or animals. But we used from the DDSM database that it has some normal texture, benign and malignant breast cancer masses.

References

1. Pawar PS, Patil DR, editors. Breast Cancer Detection Using Neural Network Models. International Conference on Communication Systems and Network Technologies; 2013 Apr 6-8; Gwalior, India: IEEE. DOI: [10.1109/CSNT.2013.122](https://doi.org/10.1109/CSNT.2013.122).
2. Rabidas R, Midya A, Chakraborty J. Neighborhood Structural Similarity Mapping for the Classification of Masses in Mammograms. IEEE J Biomed Health Inform. 2018;22(3):826-34. DOI: [10.1109/JBHI.2017.2715021](https://doi.org/10.1109/JBHI.2017.2715021) PMID: [28622679](https://pubmed.ncbi.nlm.nih.gov/28622679/).
3. Paramkusham S, Rao KMM, Prabhakar Rao BVVSN, editors. Early stage detection of breast cancer using novel image processing techniques, Matlab and Labview implementation. 15th International Conference on Advanced Computing Technologies (ICACT); 2013 Sep 21-22; Rajampet, India IEEE. DOI: [10.1109/ICACT.2013.6710511](https://doi.org/10.1109/ICACT.2013.6710511).
4. Chakraborty J, Midya A, Rabidas R. Computer-aided detection and diagnosis of mammographic masses using multi-resolution analysis of oriented tissue patterns. Expert Syst Appl. 2018;99:168-79. DOI: [10.1016/j.eswa.2018.01.010](https://doi.org/10.1016/j.eswa.2018.01.010).
5. Wang T, Cheng C, Chiu H, editors. Predicting post-treatment survivability of patients with breast cancer using Artificial Neural Network methods. 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2013 Jul 3-7; Osaka, Japan IEEE. DOI: [10.1109/EMBC.2013.6609744](https://doi.org/10.1109/EMBC.2013.6609744).
6. Abbosh YM, Yahya AF, Abbosh A, editors. Neural networks for the detection and localization of breast cancer. International Conference on Communications and Information Technology (ICCIT); 2011 Mar 29-31; Aqaba, Jordan: IEEE. DOI: [10.1109/ICCITECH-NOL.2011.5762669](https://doi.org/10.1109/ICCITECH-NOL.2011.5762669).
7. Vadivel A, Surendiran B. A fuzzy rule-based approach for characterization of mammogram masses into BI-RADS shape categories. Comput Biol Med. 2013;43(4):259-67. DOI: [10.1016/j.compbimed.2013.01.004](https://doi.org/10.1016/j.compbimed.2013.01.004) PMID: [23414779](https://pubmed.ncbi.nlm.nih.gov/23414779/).
8. Beheshti SMA, Noubari HA, Fatemizadeh E, Khalili M. Classification of abnormalities in mammograms by new asymmetric fractal features. Biocybern Biomed Eng. 2016;36(1):56-65. DOI: [10.1016/j.bbe.2015.07.002](https://doi.org/10.1016/j.bbe.2015.07.002).
9. Singh AK, Gupta B. A Novel Approach for Breast Cancer Detection and Segmentation in a Mammogram. Procedia Comput Sci. 2015;54:676-82. DOI: [10.1016/j.procs.2015.06.079](https://doi.org/10.1016/j.procs.2015.06.079).
10. Al-Shamlan H, El-Zaart A, editors. Feature extraction values for breast cancer mammography images. International Conference on Bioinformatics and Biomedical Technology; 2010 Apr 16-18; Chengdu, China: IEEE. DOI: [10.1109/ICBBT.2010.5478947](https://doi.org/10.1109/ICBBT.2010.5478947).
11. Setiawan AS, Elysia, Wesley J, Purnama Y. Mammogram Classification using Law's Texture Energy Measure and Neural Networks. Procedia Comput Sci. 2015;59:92-7. DOI: [10.1016/j.procs.2015.07.341](https://doi.org/10.1016/j.procs.2015.07.341).
12. Dhahbi S, Barhoumi W, Kurek J, Swiderski B, Kruk M, Zagrouba E. False-positive reduction in computer-aided mass detection using mammographic texture analysis and classification. Comput Methods Programs Biomed. 2018;160:75-83. DOI: [10.1016/j.cmpb.2018.03.026](https://doi.org/10.1016/j.cmpb.2018.03.026) PMID: [29728249](https://pubmed.ncbi.nlm.nih.gov/29728249/).
13. Chougrad H, Zouaki H, Alheyane O. Multi-label transfer learning for the early diagnosis of breast cancer. Neurocomputing. 2019;[Epub ahead of print]. DOI: [10.1016/j.neucom.2019.01.112](https://doi.org/10.1016/j.neucom.2019.01.112).
14. Guan S, Loew M, editors. Breast Cancer Detection Using Transfer Learning in Convolutional Neural Networks. IEEE Applied Imagery Pattern Recognition Workshop (AIPR); 2017 Oct 10-12; Washington, DC, USA: IEEE. DOI: [10.1109/AIPR.2017.8457948](https://doi.org/10.1109/AIPR.2017.8457948).
15. Chatzistergos SE, Andreadis I, Nikita KS. Identification of architectural distortions in mammograms using local binary patterns and radial lengths through an exhaustive evaluation framework. Expert Syst. 2018;35(4):e12281. DOI: [10.1111/exsy.12281](https://doi.org/10.1111/exsy.12281).
16. Digital Database for Screening Mammography USA: University of South Florida; [cited 2019 Nov 13]. Available from: <http://www.eng.usf.edu/cvprg/Mammography/Database.html>.
17. Masoodi P, Safdarian N, Kalantar B, editors. Detection and Classification of Breast Cancer using Feature Extraction from Mammography Images by Image Processing Technique and Neural Network. International Conference on Non-Linear System & Optimization in Computer & Electrical Engineering; 2015; Dubai: May 2015.
18. Jafarnia Dabanloo N, Safdarian N, Attarodi G, Afshari B, Ali S. Classification of Cardiac Arrhythmias Using Probabilistic Neural Networks. World Acad Sci Eng Technol. 2012;62:1978-81.