

Article

A Novel Computer-Aided-Diagnosis System for Breast Ultrasound Images Based on BI-RADS Categories

Yi-Wei Chang ^{1,2} , Yun-Ru Chen ³, Chien-Chuan Ko ^{1,*}, Wei-Yang Lin ³ and Keng-Pei Lin ² 

¹ Department of Computer Science and Information Engineering, National Chiayi University, Chiayi 60004, Taiwan; ywchang@g-mail.nsysu.edu.tw

² Department of Information Management, National Sun Yat-sen University, Kaohsiung 80424, Taiwan; kplin@mis.nsysu.edu.tw

³ Department of Computer Science and Information Engineering, National Chung Cheng University, Chiayi 62102, Taiwan; graces801@gmail.com (Y.-R.C.); wylin@cs.ccu.edu.tw (W.-Y.L.)

* Correspondence: kocc@mail.ncyu.edu.tw; Tel.: +886-5-2717732

Received: 10 February 2020; Accepted: 3 March 2020; Published: 6 March 2020



Abstract: The breast ultrasound is not only one of major devices for breast tissue imaging, but also one of important methods in breast tumor screening. It is non-radiative, non-invasive, harmless, simple, and low cost screening. The American College of Radiology (ACR) proposed the Breast Imaging Reporting and Data System (BI-RADS) to evaluate far more breast lesion severities compared to traditional diagnoses according to five-criterion categories of masses composition described as follows: shape, orientation, margin, echo pattern, and posterior features. However, there exist some problems, such as intensity differences and different resolutions in image acquisition among different types of ultrasound imaging modalities so that clinicians cannot always identify accurately the BI-RADS categories or disease severities. To this end, this article adopted three different brands of ultrasound scanners to fetch breast images for our experimental samples. The breast lesion was detected on the original image using preprocessing, image segmentation, etc. The breast tumor's severity was evaluated on the features of the breast lesion via our proposed classifiers according to the BI-RADS standard rather than traditional assessment on the severity; i.e., merely using benign or malignant. In this work, we mainly focused on the BI-RADS categories 2–5 after the stage of segmentation as a result of the clinical practice. Moreover, several features related to lesion severities based on the selected BI-RADS categories were introduced into three machine learning classifiers, including a Support Vector Machine (SVM), Random Forest (RF), and Convolution Neural Network (CNN) combined with feature selection to develop a multi-class assessment of breast tumor severity based on BI-RADS. Experimental results show that the proposed CAD system based on BI-RADS can obtain the identification accuracies with SVM, RF, and CNN reaching 80.00%, 77.78%, and 85.42%, respectively. We also validated the performance and adaptability of the classification using different ultrasound scanners. Results also indicate that the evaluations of F-score based on CNN can obtain measures higher than 75% (i.e., prominent adaptability) when samples were tested on various BI-RADS categories.

Keywords: Breast Ultrasound Images; CAD system; BI-RADS; SVM; RF; CNN

1. Introduction

The commonly used modalities in the diagnosis of breast carcinoma include mammography, Breast Ultrasound (BUS), Magnetic Resonance Imaging (MRI), and Computed Tomography (CT). The major advantages of BUS are it being non-radiative, non-invasive, simple, and low cost for

screening. This screening tool is suitable for women at any age, especially young women under the age of 35. Medical ultrasound transmits high-frequency sound waves via a transducer over the breast. Once the waves hit a tissue or structure, it bounces back such that the transducer receives changes of waves to create a black-and-white image of breast tissues and structures called a sonogram. Various tissues can be visualized via a sonogram under different volumes, pitches, and frequencies. This information can help a physician confirm the diagnosis that a breast lesion is identified as benign or malignant. The BUS has a panoramic view of different tissues of the breast which is composed of skin, subcutaneous, fat, muscle, glands, chest wall, and other tissues. Above all, the BUS is real-time that can obtain the result quickly. Therefore, it has become one of the major pieces of equipment for breast lesion detection—they include breast cancers, breast cysts, benign tumors, fibrocystic breast, etc. Traditionally, the severity of the patient's lesion was assessed as benign or malignant by clinicians' subjective and qualitative judgment. This cannot be effectively used to provide diagnosis or treatment for patients. In addition, it may cause some unnecessary surgery or pathological biopsy for a patient having suspicious breast carcinoma. To address this problem, the American College of Radiology (ACR) proposed a gold standard in 1993; namely, the Breast Imaging Reporting and Data System (BI-RADS). The initial edition of BI-RADS was created in 1993 and the newest edition of BI-RADS is the 5th edition [1].

BI-RADS provides a uniform standard in breast carcinoma severity and sorts the results into various categories numbered 0 through 6 according to different degrees of severity, as described in Table 1. It also provides mass descriptors examined with ultrasound, including shape, orientation, margin, echo pattern, and posterior features, as listed in Table 2. Clinically, radiologists analyzed and identified the characteristics of tumor composition intuitively and clinicians assessed the severity of breast tumors as benign or malignant according to this standard. Furthermore, this can reduce the differences between interpretations of breast cancer for physicians so that patients can obtain appropriate treatment.

Table 1. Concordance between BI-RADS assessment categories and management recommendations [1].

Assessment	Management	Likelihood of Cancer
Category 0: Incomplete—Need Additional Imaging Evaluation	Recall for additional imaging	N/A
Category 1: Negative	Routine screening	Essentially 0% likelihood of malignancy
Category 2: Benign	Routine screening	Essentially 0% likelihood of malignancy
Category 3: Probably Benign	Short-interval (6-month) follow-up or continued surveillance	> 0% but ≤ 2% likelihood of malignancy
Category 4: Suspicious	Tissue diagnosis	> 2% but ≤ 95% likelihood of malignancy
Category 4A: Low suspicion for malignancy	Tissue diagnosis	> 2% to ≤ 10% likelihood of malignancy
Category 4B: Moderate suspicion for malignancy	Tissue diagnosis	> 10% to ≤ 50% likelihood of malignancy
Category 4C: High suspicion for malignancy	Tissue diagnosis	> 50% to ≤ 95% likelihood of malignancy
Category 5: Highly Suggestive of Malignancy	Tissue diagnosis	≥ 95% likelihood of malignancy
Category 6: Known Biopsy-Proven Malignancy	Surgical excision when clinically appropriate	N/A

In Table 2, the shape category represents the shape of mass including: round, oval, or irregular. Orientation refers to the lesion’s long axis relative to the skin; the descriptors for the mass margin are circumscribed or non-circumscribed, which in non-circumscribed margins include micro-lobulated, indistinct, angular, and speculated; the descriptors for the mass echo pattern rely on tissue having various reference echotextures; the posterior feature refers to the attenuation characteristics of a mass relative to its acoustic transmission. In other words, BI-RADS provides a uniform standard for the interpretation of a breast image regarding the breast lesion—whether there exists a tumor; and evaluating the shape regularity of the mass, the distinction degrees of the non-circumscribed, the variabilities of the reference echotexture, the variabilities of the posterior feature, etc. (See Figure 1).

Table 2. Masses features category [1].

ULTRASOUND		
Category	Descriptors	Description
Shape	Oval	Elliptical or egg-shaped.
	Round	Spherical, ball-shaped, circular.
	Irregular	Neither round or oval in shape.
Orientation	Parallel	Long axis of lesion parallels the skin line.
	Not parallel	Long axis, not orientation along the skin line.
Margin	Circumscribed	A margin that is well defined or sharp, with an abrupt transition between the lesion and surrounding tissue.
	Not circumscribed	The mass has one or more of the following features: indistinct, angular, microlobulated or speculated.
	- Indistinct	No clear demarcation between a mass and its surround tissue.
	- Angular	Some or all of the margin has sharp corners, often forming acute angles.
	- Microlobulated	Short cycle undulations impart a scalloped appearance to the margin of the mass.
Echo pattern	- Spiculated	Margin is formed or characterized by sharp lines projecting from the mass.
	Anechoic	Without internal echoes.
	Hyperechoic	Having increased echogenicity relative to fat or equal to fibroglandular tissue.
	Complex cystic and solid	Mass contains both anechoic and echogenic components.
	Hypoechoic	Defined relative to fat; masses are characterized by low-level echoes.
	Isoechoic	Having the same echogenicity as fat.
Posterior features	Heterogeneous	The breasts are heterogeneously dense, which may obscure small masses
	No posterior features	No posterior shadowing or enhancement.
	Enhancement	Increased posterior echoes.
	Shadowing	Decreased posterior echoes.
	Combined pattern	More than on pattern of posterior attenuation, both shadowing and enhancement.

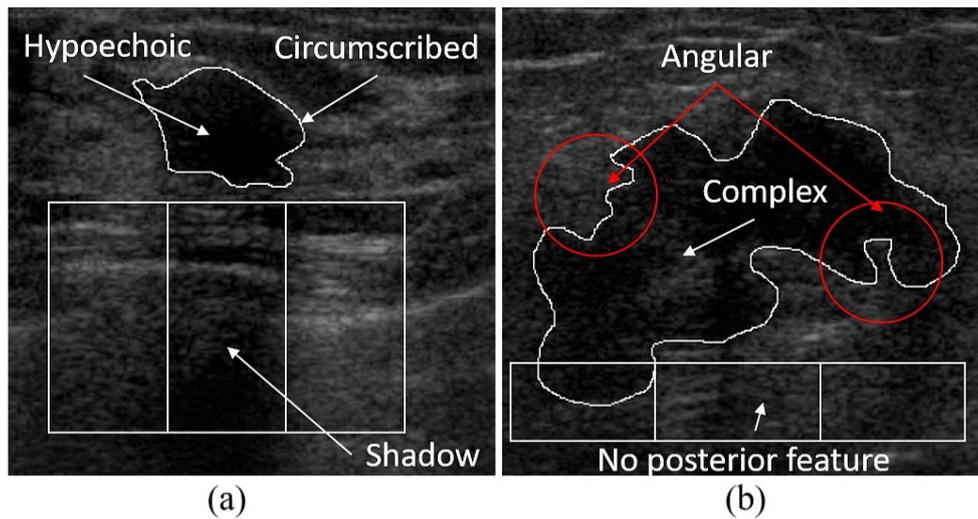


Figure 1. Illustration of BI-RADS descriptor. (a) Benign lesion (Category 3); (b) malignant tumor (Category 4).

In this paper, machine learning classifiers were applied to identify the severities of breast lesion on BUS images so as to develop an automated and accurate CAD system. The proposed CAD system can provide diagnostic references for surgeons in interpreting BI-RADS categories. Basically, the proposed CAD system can be roughly divided into the following steps: image preprocessing, image segmentation, feature extraction, feature selection and classification. Figure 2 shows a flowchart of the BI-RADS category system. Especially, this study also focuses on multi-class identification based on the BI-RADS category 2 to 5, and cope with problems such as different resolutions and intensity variations in image acquisition among different types of ultrasound imaging modalities. The rest of this paper is organized as follows. In Section II, we survey the related works on the classification of breast ultrasound BI-RADS and review the relevant techniques. Section III describes the proposed method for multi-class Computer-Aided-Diagnosis system. The experimental results are illustrated in Section IV, and Section V concludes this paper.

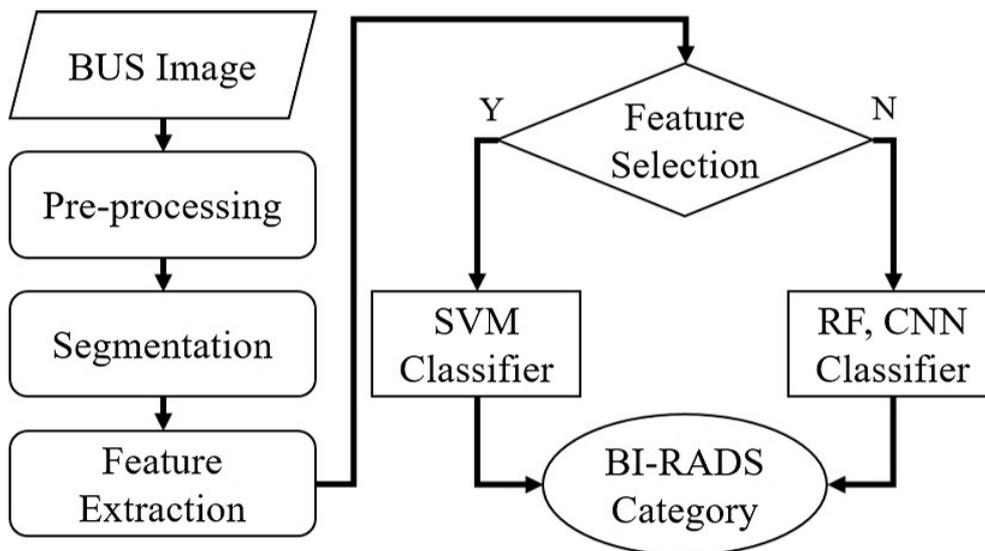


Figure 2. Flowchart of the BI-RADS category identification system.

2. Related Work

Various Computer-Aided-Diagnosis (CAD) systems in the breast imaging modality has been widely developed. In general, traditional CAD systems consist of two major stages: one is segmentation of masses, and the other is severity classification of masses. Some works also explored segmentation of breast masses. For example, Alvarenga et al. [2] used semi-automatic CAD system to detect the contours of breast lesion. Shan [3] developed a fully automatic detection system for breast lesion segmentation. Both accurate lesion boundary detection and feature selection are important for breast cancer diagnosis. The research in [2] identifies lesion severity on ultrasound images using morphology features, and the researches in [3–5] adopt texture features for classification. Both of morphology features and texture features are used for classification [6–8]. Moon et al [7] cooperates with pathological diagnosis to identify a breast lesion with BI-RADS as benign or malignant. As a whole, most of the researches only can predict the detected breast lesion tumor as benign and malignant [2–8]. In recent years, most of researches who conduct classification of breast ultrasound BI-RADS only focused on borderline category such as BI-RADS 3 to reduce the ambiguities. Related studies and their adopted methods are illustrated in Table 3, respectively.

In actuality, there exist several speckle noises on BUS images as a result of human factors or mechanical imaging factors. Traditional noise removal method cannot reduce the speckle noises effectively. Therefore, Yu and Acton [9] developed a nonlinear anisotropic diffusion technique, speckle reducing anisotropic diffusion (SRAD) to remove the speckle. Besides, how to extract robust features related to the BI-RADS description is an important issue for accurate classification. Shan [3] adopted region growing from the seed point, speckle reduction, and the Neutrosophic L-Means clustering (NLM) to develop a fully automatic segmentation method for BUS images. Although the segmentation accuracy reaches 94%, this method has two major limitations. The first one is over-segmentation problem occurred where the lesion position near the image boundary. Another problem is that it cannot detect multiple lesions in one image. In this work, we propose a method to improve the problems mentioned above.

In recent years, some studies usually capitalized on the machine learning such as the selected features and Support Vector Machine (SVM) to predict the lesion severity with a two-category classification. Recently, Random Forest (RF) has become a commonly used method for multi-class identification especially in surface-types analysis of satellite imagery [10]. Since 2012, Krizhevsky [11] proposed Convolution Neural Network (CNN) on color images to identify objects of interest. Classification researches in medical images have gradually aimed at the approaches based on deep learning. In recent years, CNN has been applied and demonstrated to tissue identification of brain MRI images effectively [12,13], because it can achieve better classification results. Nowadays, Yap et al. [14] using CNN for breast lesions detection; Chiang et al. [15] adopted 3-D CNN for breast tumor detection.

After further exploration, we ensemble the proposed methods and existing techniques into our proposed breast ultrasound CAD system so that our method can extract more robust features. The features selected by the proposed system are further illustrated in Table 4.

Table 3. Literature reviews of breast ultrasound CAD system.

Study	Year	Method	Classifier	Accuracy(F-measure)	Category
Alvarenga [2]	2010	GLCM	FLDA	83%	M,B
Shan [3]	2011	SRAD, RG	NLM	94%	Segmentation
Wu [6]	2012	PCA+GA	SVM+RBF	92%	M,B
Gómez [4]	2012	GLCM	FLDA+.632	87%	C,B
Yang [5]	2013	GLCM	SVM+.632	87%	M,B
Moon [7]	2013	Backward	BLR	88%	M,B
Shan [8]	2016	SRAD, RG	DT, ANN, RF, SVM	77.7%, 78.1%, 78.5%, 77.7%	M,B
Yap [14]	2018	HOG	LeNet, U-Net, FCN-AlexNet	(91%), (89%), (92%)	M,B
Acharya [16]	2019	LSDA, MR8	DT, k-NN, PNN, SVM	89.3%,92.3%, 96.1%, 95.3%	M,B

Notes. C: Carcinoma; B: Benign; M: Malignant; .632: 0.632+ bootstraps; RG: Region Growing; DT: Decision Tree; NLM: Neutrosophic L-Means; LDA: Linear Discriminant Analysis; BLR: Binary Logistic Regression; FLDA: Fisher Linear Discriminant Analysis; GLCM: Gray-Level Co-occurrence; LSDA: Locality Sensitive Discriminant Analysis; HOG: Histogram of Oriented Gradients; PNN: Probabilistic Neural Network.

Table 4. BI-RADS features corresponding list.

BI-RADS Categories		Features	Total	
morphology	shape	(F1) Circularity (F2) Form Factor (F3) Roundness (F4) Long axis to Short axis Ratio (F5) Aspect Ratio (F6) Elliptic-Normalized Circumference (F7) Convexity (F8) Standard Deviation of Normalized Radial Length (F9) Entropy of Normalized Radial Length (F10) Extent	10	
		Orientation	(F11) Orientation Angle	1
		Margin	(F12) Roughness Area Ratio (F13) Margin Roughness1 (F14) Solidity (F15) Normalized Residual Value (F16) Number of Substantial Protuberances and Depressions (F17) Indistinction	6
Textural	Echo Pattern	(F18) Multi-resolution Wavelet Transform + GLCM (F19) Multi-resolution Ranklet Transform + GLCM (F20) ROI + GLCM	126	
	Posterior Features	(F21) Difference of the average of gray level that between the upper and lower area (F22) Difference of the average of gray level that between the right and left area	2	

Notes. GLCM represent Gray-Level Co-Occurrence Matrix.

3. Materials

In this paper, we collected retrospective data (between 2012 and 2014) from our cooperating hospital. There was a total of 151 tumor lesions samples used to validate the performance of the proposed system. These images included 151 tumor lesions composed of 79 benign tumors (BIRADS 2–3) and 72 malignant tumors (BIRADS 4–5) and were first identified by an experienced physician. The links with personal information of these patients from these images will be further removed by an experienced radiologist. Then, each fetched ultrasound image of the patients will be stored as JPEG format with its corresponding serial number with a spatial resolution of pixels and 256 intensity values. Because the original images fetched by different imaging instruments do not have the same size, the border of each image may be directly cropped to a specified size so that some details of the acquired image may be preserved, and the size may be consistent (321×321). The original images acquired from three different imaging instruments including PHILIPS, SIEMENS, and TOSHIBA after the crop operation are shown in Figure 3, respectively. Hereinafter referred to as Model B, B, and C. We find that these images both may have different performances in various breast tissues.

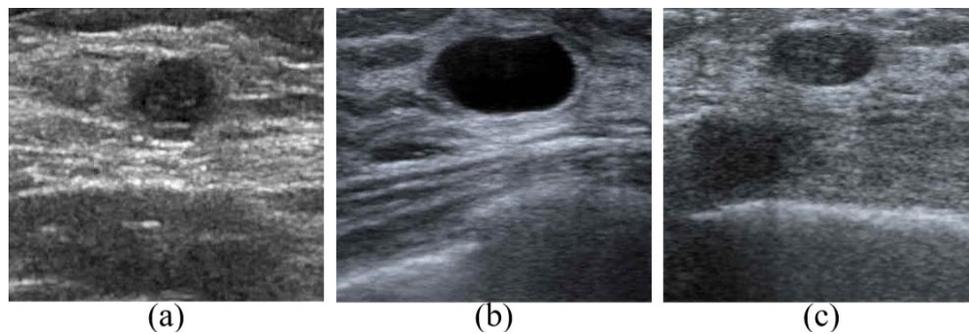


Figure 3. Different types of ultrasound imaging modalities; images compressed to the same size: (a) with Model B; (b) with Model B; (c) with Model C.

4. Methods

The major goal of the proposed CAD system is to predict the severity of breast carcinoma, and can be roughly divided into the following stages: preprocessing, segmentation/detection of breast tumor, feature extraction, and BI-RADS category prediction. Our proposed CAD system first acquires the BUS image, performs image preprocessing that include speckle noise removal, image normalization, and image enhancement, and then perform image segmentation that include k-means clustering, boundary region removal, regions ranking, and region growing. Figure 4. shows the detailed flowchart of the tumor contour detection. The individual processing results using these processing operations are illustrated in Figure 5, respectively.

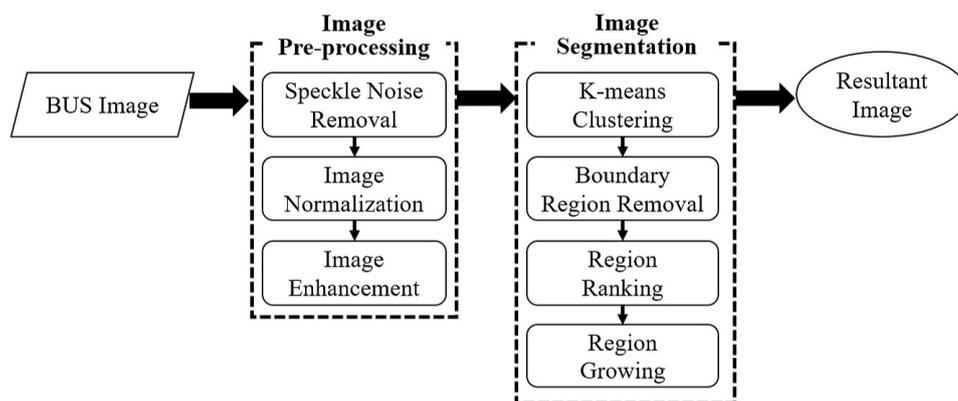


Figure 4. The flowchart for the contour detection of breast tumor.

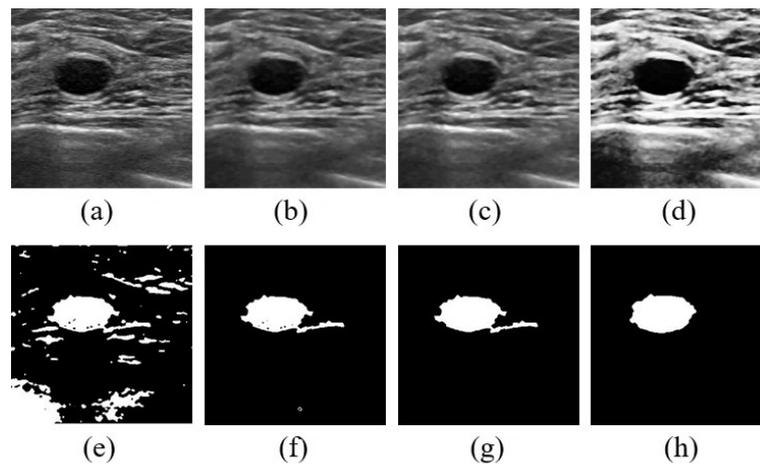


Figure 5. Contour detections for the sample tested for BI-RADS 2 category with Model B: (a) original; (b) SRAD; (c) image normalization; (d) image enhancement; (e) k-means clustering; (f) result after removing regions with small areas or regions touched with image border; (g) result image after region ranking; (h) result image after region growing.

4.1. Image Preprocessing

In the preprocessing stage, speckle noise reduction based on anisotropic diffusion (SRAD) [17] was conducted to remove the speckle noise caused by different diffusion signals. It reduced the interference of noises on the breast tissues, as shown in Figure 5b. Next, intensity normalization was applied to decrease the brightness differences among all images acquired by different types of imaging instruments. This operation preserves image quality, and eliminates variabilities from different radiologists who fetch images with different ultrasound machines, as shown in Figure 5c. Finally, a contrast enhancement method based on histogram equalization was performed to increase the contrast between the tumor regions and the non-tumor (background) regions as shown in Figure 5d.

4.2. Segmentation of Breast Tumor

In the segmentation stage, we only focus on how to detect effectively true tumor regions for latter processing such as feature extraction, and classification. To this end, the tumor Region Of Interest (ROI) needs to be located beforehand. An initial segmentation based on K-means clustering [18] was performed to detect the suspicious tumor areas including tumor and background which is composed of fat, skin, muscle, glands and other tissues, as shown in Figure 5e. Since the dark tissues around a tumor are easy to be mis-classified as candidate tumors after K-means clustering, eliminating the pixels which do not belong to the tumor tissue using some processing operations is necessary in order to improve the clustering accuracies. In this step, operation based on morphological dilation reconstruction was applied to remove the regions touching the border of the image, and objects with smaller area were considered as artifacts and further filtered out. The remaining regions may thus be considered as suspicious/candidate tumor or lesion regions, as shown in Figure 5f.

Unfortunately, after the previous step, some suspicious tumor regions may be preserved yet. Clinically, a tumor may be found as an object having a round shape. In order to detect an actual tumor region, a region ranking method [3] based on scoring in terms of area similarity and circularity was applied on Figure 5f to determine the region whose ranking score was greater than a specified threshold. Thus, candidate tumor areas may be further selected to be tumor areas, as shown in Figure 5g. However, there exist holes or irregular shapes in the detected tumor regions due to non-uniform intensities caused by different skills or experience in fetching images. Seed regions inside these candidate regions may be determined and used as reference points for region growing [19]. In order to resolve this problem, we used the growing method proposed by Shan [3]. The seed point is used as a start point,

and a region will be expanded gradually by observing its neighborhood pixels' intensity values and gradient magnitudes. This operation can be expressed as Equation (1):

$$G(v) \leq \max\left(\frac{M}{b_2}, \min(b_1 \times m, M)\right) \quad (1)$$

where $G(v)$ is the intensity value of pixel v , m is the intensity mean of region, and M is the intensity mean of the entire image. In our experiment, we used $b_1 = 1$ and $b_2 = 1.9$ as our parameters. Nevertheless, the proposed approach still cannot accurately detect the tumor acquired by difference ultrasound machines. Therefore, the growing method is further modified from Equation (1) to Equation (2) in order to resolve the problem coming from different machines:

$$G(v) \leq \frac{(M + m)}{2} - b \quad (2)$$

where $G(v)$ is the intensity value in pixel v , m is the average intensity of the growing area, M is the average intensity of the entire image, and b is criterion parameter for halted growing. According to the different sonogram models, the average parameter of b is determined by various experiments which are set as 7, 18, and 14 for Model B, Model B, and Model C, respectively. After performing the region growing, an actual breast tumor can thus be detected, as shown in Figure 5h. Resultant images of the detected tumor contour with different machines are shown in Figure 6, respectively. Therefore, the proposed breast ultrasound CAD system not only can segment multi-tumors, but also can be suitable for different types of breast ultrasonic scanners. Especially, it even can detect multiple tumors automatically.

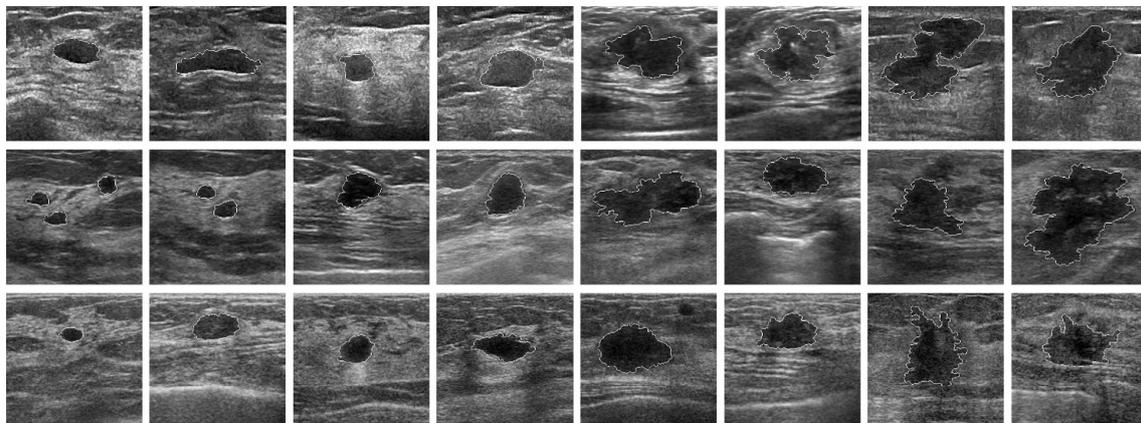


Figure 6. Resultant images after tumor contour detection with different machines. First row is the resultant images tested with Model B, second row is the resultant images tested with Model B images, and third row is resultant images tested with Model C; images in column one and two are results tested with BI-RADS 2 cases; images in column three and four are results tested with BI-RADS 3 cases, images in column five and six are results tested with BI-RADS 4 cases; images in column seven and eight are results tested with BI-RADS 5.

4.3. Feature Extraction

Once the actual tumor contour has been located, how to extract robust features is important for correct BI-RADS classification. Here, morphology features and texture features which correspond to the BI-RADS descriptors are defined and measured. The flow chart of the feature extraction stage is illustrated in Figure 7. A total of 145 features composed of their corresponding descriptors, including shape features, orientation features, margin features, echo pattern, posterior features, etc., as mentioned in the Table 4, were extracted to identify the BI-RADS category. These 17 features corresponding to morphology, such as shape, orientation, and margin, were measured quantitatively.

Besides, texture can be efficiently used to describe echo pattern and the posterior, as mentioned in Table 4. This is due to the fact that it can measure these two feature categories using the spatial intensity distribution of the tissues or using degree of degradation when the ultrasound signals penetrate the tissues. In general, gray level co-occurrence matrix (GLCM) [4] based on orientation decomposition is a good measure for texture feature, and is evaluated often by distances and orientation between neighboring pixels. Here, GLCM at four directions (0° , 45° , 90° , 135°) and one distance ($d = 1$ pixel) are first constructed; G is referred to as the matrix. Then, an estimate of the probability p_{ij} is computed from the elements of G . A variety of texture descriptors used for characterizing the contents of GLCM, including contrast, correlation, energy, entropy, homogeneity, sum average, and sum entropy, were computed and averaged, respectively. Therefore, a total of seven features were chosen when we performed GLCM on the original image. Besides, multi-resolution wavelet transform (adopting eight images) and multi-resolution ranklet transform (adopting nine images) were also performed using the same descriptors measured from the same operations by using GLCM, respectively. Finally, a total of 126 features were generated from the original image, the multi-resolution wavelet transform, and the multi-resolution ranklet transform. Because the posterior feature can be used to evaluate the echo variations behind the tumor, it can be often measured directly from the intensity distributions close to the tumor region via computing the difference of the average between the upper and lower area, and the difference of the average between the right and the left area (generating two features). Finally, a total of 128 mass features corresponding to BI-RADS descriptors such as echo pattern and posterior features were extracted in this stage [4]. However, only the SVM classifier needs to be performed using a feature selection stage, such as feature dimension reduction, and feature scaling, in order to exclude unnecessary or insignificant features, because the classifier itself does not have the capability of feature selection.

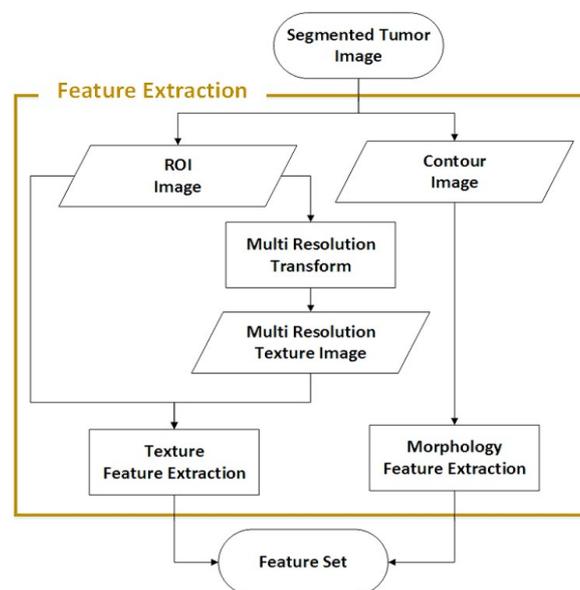


Figure 7. The flowchart of the BI-RADS feature extraction stage.

Various features related to lesion or disease severities corresponding to the BI-RADS categories were fed into three machine learning and deep learning classifiers, including: Support Vector Machine (SVM), Random Forest (RF), and Convolution Neural Network (CNN) with/without feature selection operation to develop a multi-class assessment of breast tumor severity based on BI-RADS, as shown in Figure 8. From these three classifiers, a classifier evaluated with the highest performance was chosen as our final classification decision for the proposed CAD system.

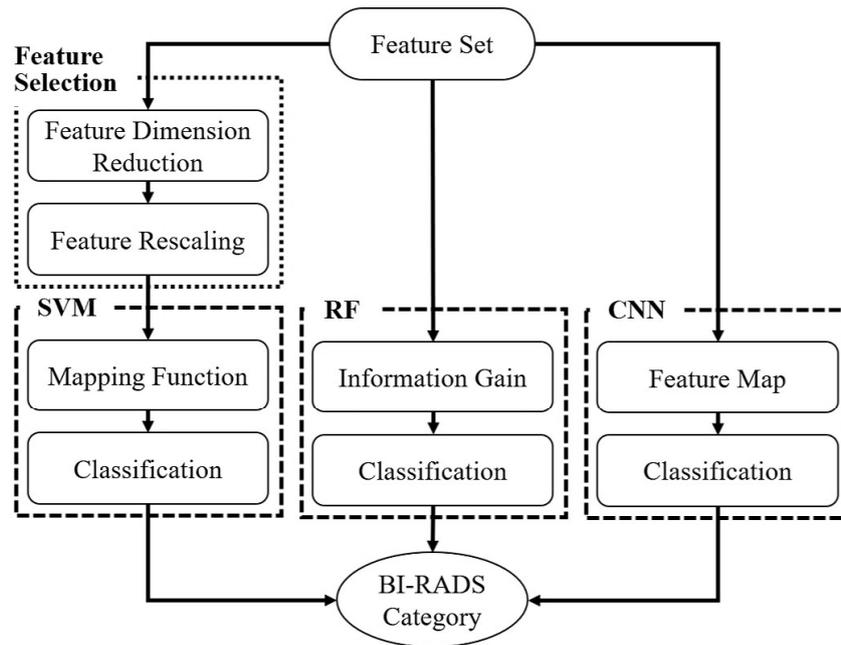


Figure 8. The overview of the process of feature selection and classifier.

4.4. Classifiers

Three types of machine learning techniques were applied to predict the BI-RADS category, including SVM, RF, and CNN. The implementations of these classifiers are described below.

Support Vector Machine

Basically, SVM is a two-class classifier based on statistical theory in the field of machine learning. In this study: a multi-class SVM classifier with Radial Basis Function (RBF) kernel function whose gamma function and penalty coefficient C were adjusted continuously in order to obtain better performance and speed up the entire computation process. LIBSVM [20] utilized the model, and was implemented on the Matlab. The RBF (Gaussian) kernel function was used as a mapping transform function for SVM. The model calculates the correction rate after mapping and compares the correction rate to obtain a better trade-Off. Finally, it obtains the optimal plane parameter set, which is $C = 2^{x_i}$, $gamma = 2^{x_j}$. Experimental testing results show that optimal parameter C and gamma value are set as 181.0193 and 0.088388 in terms of our experiments, respectively. The proposed classifier would obtain best classification performance. The stage aims to design a multi-class SVM classifier [21] from a number of optimal planes in a large data space. In fact, a multi-classification between different categories not only requires a lot of computation time, but will reduce identification accuracy indirectly. SVM often combines the Heuristic algorithm to evaluate the robustness of features correlation and selectivity in a reasonable time. In order to determine the best solution, it is necessary to avoid this problem using more resources on computation time. In this paper, a multi-class SVM based on a Directed Acyclic Graph (DAG-SVM) [22] was presented to predict the BI-RADS category, and the BI-RADS classification architecture of the BI-RADS DAG-SVM is shown in Figure 9.

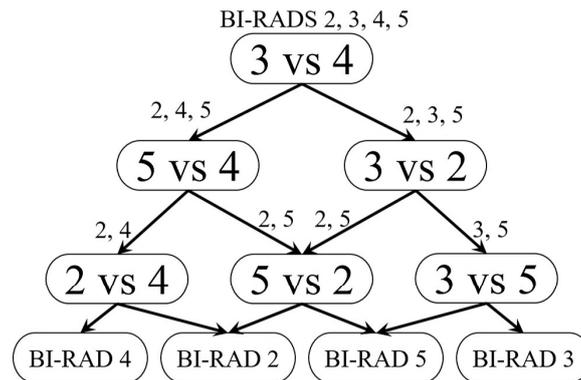


Figure 9. Diagram of the proposed DAG-SVM architecture.

4.5. Random Forest

A Decision Tree is one of the commonly used predictive models in machine learning. It is a non-parametric supervised learning method used for both classification and regression tasks, whereas a Random Forest (RF) [23] is generated randomly by various Decision Trees (DT). Above all, it uses majority voting strategy to obtain the best prediction results from preliminary classification results. As shown in Figure 10, the optimal subset of each tree can be transformed into a variable from the feature set x , and a total of Decision Trees N is thus established. The optimal decision of each tree is transformed to a variable; i.e., the overall classification result Y . That is, RF can automatically select the best combination from individual trees based on gain ratio and split information. The classifier RF can be denoted as Equation (3) [24]:

$$\hat{Y}(x) = \arg \max_c \frac{1}{N} \sum_{t=1}^N P_t(Y(x) = c) \tag{3}$$

where c is the class label, and P_t is the BI-RADS category probability measured by DT. After evaluating the performance of the forest with mean square error and tree number equal 200, out of bag (OOB) error curve will not be significantly reduced, and then the classification may reach a steady state finally. This operation not only minimize error measure, but ensure that the proposed prediction Model Can obtain better identification accuracies between the training stage and the testing stage as much as possible. Our experiments also show that if the tree number is greater than 200, the RF classification efficiency no longer increase significantly. Therefore, this stage selects 200 trees as a limitation criterion for our RF classifier.

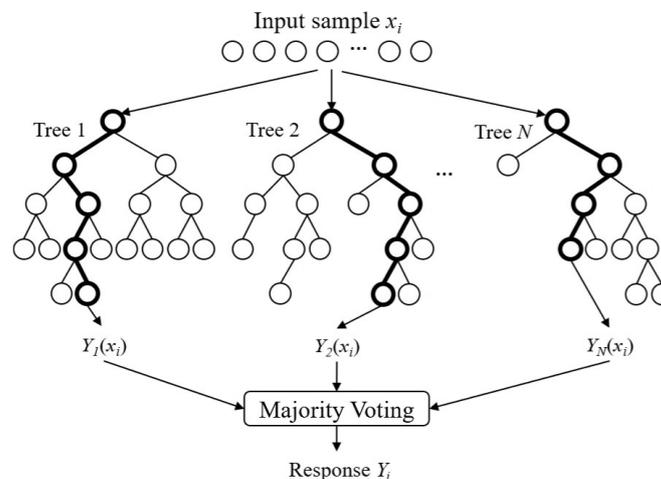


Figure 10. Diagram of RF architecture.

4.6. Convolutional Neural Network

Convolutional Neural Network (CNN) [11]: a class of deep artificial neural networks; it has become a popular classification approach in computer vision recently. It consists of an input layer, a convolution layer, a pooling layer, a fully connected layer, and an output layer. A feature map is obtained by processing the input image with the convolutional kernel; weight-sharing is used among neurons in the same convolutional layer. During back propagation, convolution layers get updated to let the network learn the features itself. The weights or bias for every neuron may be modified or updated using gradient descent and backpropagation where the chain rule of calculus is used to adjust the weights that will minimize the loss function.

In this study, a CNN architecture was implemented on the BUS image in order to validate the classification performance of the model. Before each acquired BUS image was introduced into the model, the size of the image was resized to a spatial resolution of 99×99 pixels in order to satisfy the input criterion of the proposed CNN model. Because the sample size of the data set is restricted, effective data preprocessing and augmentation is mandatory for medical image datasets, especially for CNN training. In the CNN training stage, only geometric translation and rotation were performed to preserve the shape textures of breast tumors for the final classification. Moreover, all BUS image samples were decomposed into five categories, including background itself (negative sample) and BI-RAD categories 2–5 (positive sample). The illustration of CNN on an input image is shown in Figure 11. The CNN architecture includes an input layer, two convolution layers, and two max-pooling layers, and two fully connected layers for the multi-class classification are also shown in Figure 12. The fully connected layer 1 flattens max-pooling layer 2 into a single vector. The detailed parameters of the proposed CNN architecture are further summarized in Table 5.

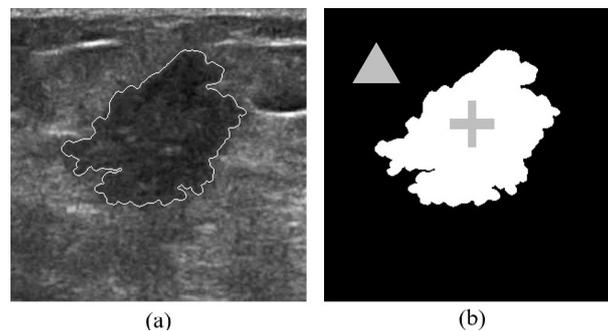


Figure 11. Illustration of CNN on an input image: (a) original image with a size of 321×321 ; (b) sample image cropped on the original image—size was reduced to 99×99 . The cross-shaped is a positive sample (tumor), but the triangular one is a negative sample (background).

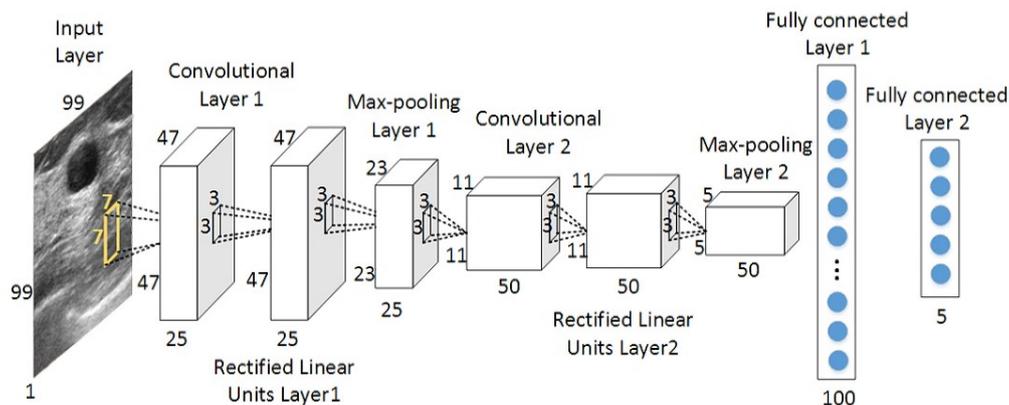


Figure 12. The proposed CNN architecture.

Table 5. Details of the proposed CNN layer detail.

Layer	Type	Input Size	Detail
Layer 0	Input	99×99	1 image
Layer 1	Convolution	47×47	$K = 7 \times 7, N = 25$
Layer 2	Rectified Linear Unit	47×47	$K = 3 \times 3, N = 25$
Layer 3	Max Pooling	23×23	$K = 3 \times 3, N = 25$
Layer 4	Convolution	11×11	$K = 3 \times 3, N = 50$
Layer 5	Rectified Linear Unit	11×11	$K = 3 \times 3, N = 50$
Layer 6	Max Pooling	5×5	$N = 50$
Layer 7	Fully connected	100×1	100 node
Layer 8	Fully connected	5×1	5 class

K = mask size = kernel size; N = number of mask.

5. Experimental Results and Discussion

In our experiments, we collected retrospective data from our cooperating hospital in 2012. There is a total of 151 tumor lesion samples. These 151 images were decomposed into 106 training samples and 45 testing samples for evaluating the performance of SVM. In order to evaluate the performance of RF and CNN, 103 training samples and 48 testing samples were used.

5.1. Performance Evaluation with Confusion Matrix

As mentioned previously, in order to validate the detection performance of the proposed CAD, we compared the segmentation results with Equation (1). Figure 13 shows the contour detection results using different approaches on different ultrasound instruments, including: Model B, Model B, and Model C.

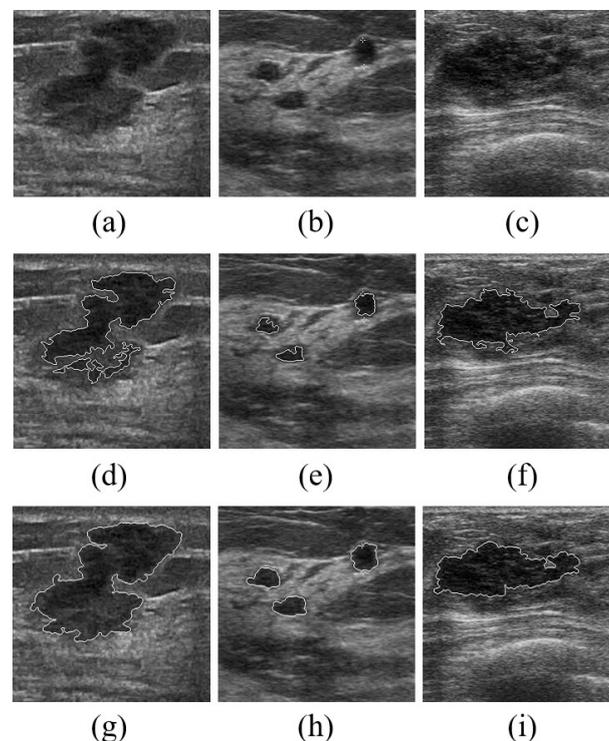


Figure 13. Performance comparison of the contour detection with the proposed CAD. The first row lists original images; second row lists the processing results with the method defined in Equation (1); and third row lists the contour detection with the proposed CAD system. (a,d,g) Images fetched with Model B; (b,e,h) images fetched with Model B; (c,f,i) images fetched with Model C.

5.2. Performance Evaluation with Confusion Matrix

In general, confusion matrix is a commonly used index to evaluate the efficiency of the machine learning. In order to validate the performance of the multi-class identification, the predicted and actual BI-RADS categories (ground truth) were compared to generate a confusion matrix [25]. As shown in Figure 14, a 2D confusion matrix is generated in terms of the predicted and actual classification results using different machine learning approaches. Assume that K_1, K_2, \dots, K_3 are the multiclass label; A_{ij} denotes that the actual class is K_i for each sample, but the predicted class is K_j . According to the confusion matrix, various indices for the performance evaluation can be measured so as to evaluate the classification efficiency. Some popular indicators will be expressed further in terms of the following definitions.

Accuracy is used to evaluate the correction percentage for all categories, and the measure is calculated by Equation (4):

$$Accuracy = \frac{\sum_{i=1}^n A_{ii}}{\sum_{i=1}^n \sum_{j=1}^n A_{ij}} \quad (4)$$

Precision is its prediction accuracy for a specific class. In other words, it is the measurement accuracy of successful prediction for a sample, and it is calculated by Equation (5):

$$Precision_i = \frac{A_{ii}}{\sum_{k=1}^n A_{ki}} \quad (5)$$

Recall is an accuracy measure of the prediction model (the test model of the classifier) in the measurement of a particular class. In other words, the probability for a certain category is not misclassified, and its definition is denoted as Equation (6):

$$Recall_i = \frac{A_{ii}}{\sum_{k=1}^n A_{ik}} \quad (6)$$

The statistical *F-score* (*F1-score*) is the harmonic mean calculated by the precision rate and the recall rate. When the *F-score* is over 70%, the test method will be considered to be more effective, and the measure is calculated by Equation (7):

$$F-score_i = \frac{2 \times Precision_i \times Recall_i}{Precision_i + Recall_i} \quad (7)$$

In our experiments, there are a total of 106 training samples and 45 testing samples which were selected and identified by an experienced radiologist. The evaluation using SVM+PCA to classify the BI-RADS category is summarized in Table 6. Besides, the evaluation using RF to classify the BI-RADS category is also shown in Table 7. In order to evaluate the performance of CNN, 103 training samples and 48 testing samples were used. The evaluation result for the method is shown in Table 8. In terms of three confusion matrices, Tables 6–8, the classification efficiency can be evaluated and compared, as shown in Table 9. Actually, optimal performance can be obtained using CNN for various evaluation indices in different BI-BARDS category classification.

		Predicted		
		K_1	$\dots K_j \dots$	K_n
Actual	K_1	A_{11}	A_{1j}	A_{1n}
	\vdots		\vdots	
	K_i	A_{i1}	$\dots A_{ij} \dots$	A_{in}
	\vdots		\vdots	
	K_N	A_{n1}	A_{nj}	A_{nn}

Figure 14. Illustration of confusion matrix.

Table 6. SVM confusion matrix.

		Predicted			
		BI-RADS 2	BI-RADS 3	BI-RADS 4	BI-RADS 5
Actual	BI-RADS 2	20	0	0	2
	BI-RADS 3	2	1	0	0
	BI-RADS 4	2	1	2	2
	BI-RADS 5	0	0	0	13

Table 7. RF confusion matrix.

		Predicted			
		BI-RADS 2	BI-RADS 3	BI-RADS 4	BI-RADS 5
Actual	BI-RADS 2	17	1	1	0
	BI-RADS 3	1	3	1	0
	BI-RADS 4	0	4	4	1
	BI-RADS 5	0	0	1	11

Table 8. CNN confusion matrix

		Predicted			
		BI-RADS 2	BI-RADS 3	BI-RADS 4	BI-RADS 5
Actual	BI-RADS 2	16	1	0	0
	BI-RADS 3	2	7	0	0
	BI-RADS 4	0	0	6	3
	BI-RADS 5	0	0	1	12

Table 9. A comparison of the performances of different machine learning classifiers.

Classifier	BI-RADS	Recall (%)	Precision (%)	F-score	Accuracy (%)
SVM	2	90.91 %	83.33%	86.96%	80.00%
	3	33.33 %	50.00%	40.00%	
	4	28.57 %	100.00%	44.44%	
	5	100.00%	76.47%	86.67%	
RF	2	89.47 %	94.44%	91.89%	77.78%
	3	60.00 %	37.50%	46.15%	
	4	44.44 %	57.14%	50.00%	
	5	91.67 %	91.67%	91.67%	
CNN	2	94.12 %	88.89%	91.43%	85.42%
	3	77.78 %	87.50%	82.35%	
	4	66.67 %	85.71%	75.00%	
	5	92.31 %	80.00%	85.71%	

6. Conclusions

In the study, three types of ultrasound imaging instruments including Model B, Model B, and Model C were adopted to acquire breast ultrasound images for our experimental samples. A series of image processing operations such as speckle noise removal, image normalization, and image enhancement etc. were performed to detect the breast lesion contours, extract features related to BI-RADS category with/without feature selection procedure depending on the selected classifier. Finally, the multi-class classifiers based on machine learning methods were implemented to identify actual BI-RADS 2-5 Categories in terms of the selected features. Experimental results also reveal that the proposed system not only can automatically, accurately, and reliably detect multi-tumors, but also show that the identification accuracies performed with SVM, RF, and CNN were evaluated as 80.00%, 77.78%, and 85.42%, respectively. CNN is higher in accuracy than RF and SVM. Because CNN model used is a pixel-based structure with feature map in extracting features whose features can be more than a total of 145 mass features that were extracted in our proposed feature selection. Using different ultrasound imaging instruments, evaluations of F-score based on CNN can obtain various measures higher than 75% when samples were tested on various BI-RADS categories.

In the future, we hope to acquire sufficient BI-RADS 3–4 samples from ultrasound elastography and B-Mode ultrasound images in order to improve and validate the evaluation performance. Cross-validation can thus also be demonstrated if we can further combine with much more ultrasound samples and histopathological examination results so that the performance of the feature extraction and the classification can be further improved in the proposed CAD system.

Author Contributions: C.-C.K. envisioned and coordinated the invited paper preparation; Y.-R.C. contributed to manuscript preparation, most of the experiments, and partial writing; Y.-W.C. and K.-P.L. contributed to manuscript writing and paper revision including some of the experiments and experimental analysis. C.-C.K. and W.-Y.L. supervised the process. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: This research is partially supported by the Ministry of Science and Technology, Taiwan, R.O.C. under Grant no. MOST 105-2221-E-415-020-MY2. Si-Wa Chan identified BI-RADS category on the selected breast ultrasound images.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ACR	American College of Radiology
BI-RADS	Breast Imaging Reporting and Data System
CAD	Computer-Aided-Diagnosis
SVM	Support Vector Machine
RF	Random Forest
CNN	Convolution Neural Network
GLCM	Gray-Level Co-occurrence
FLDA	Fisher Linear Discriminant
C	Carcinoma
B	Benign
M	Malignant
SRAD	Speckle Reducing Anisotropic Diffusion
NLM	Neutrosophic L-Means
PCA	Principal Component Analysis
GA	Genetic Algorithms
RBF	Radial Basis Function
FLDA	Fisher Linear Discriminant Analysis
BLR	Binary Logistic Regression
RG	Region Growing
ANN	Artificial neural networks
HOG	Histogram of Oriented Gradients
MR8	Maximum Response (eight responses)
K-NN	K Nearest Neighbor
PNN	Probabilistic Neural Network
ROI	Region Of Interest
LIBSVM	A Library for Support Vector Machines
DAG-SVM	Directed Acyclic Graph–Support Vector Machines

References

1. D’Orsi, C.J.; Sickles, E.A. *ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System*, 5th ed.; American College of Radiology: Reston, VA, USA, 2013.
2. Alvarenga, A.V.; Infantosi, A.F.C.; Pereira, W.C.A.; Azevedo, C.M. Assessing the performance of morphological parameters in distinguishing breast tumors on ultrasound images. *Med. Eng. Phys.* **2010**, *32*, 49–56, doi:10.1016/j.medengphy.2009.10.007.
3. Shan, J. A Fully Automatic Segmentation Method for Breast Ultrasound Images. Ph.D. Thesis, Utah State University, Logan, UT, USA, 2011.
4. Gomez, W.; Pereira, W.C.A.; Infantosi, A.F.C. Analysis of Co-Occurrence Texture Statistics as a Function of Gray-Level Quantization for Classifying Breast Ultrasound. *IEEE Trans. Med. Imaging* **2012**, *31*, 1889–1899, doi:10.1109/TMI.2012.2206398.
5. Yang, M.; Moon, W.K.; Wang, Y.F.; Bae, M.S.; Huang, C.; Chen, J.; Chang, R. Robust Texture Analysis Using Multi-Resolution Gray-Scale Invariant Features for Breast Sonographic Tumor Diagnosis. *IEEE Trans. Med. Imaging* **2013**, *32*, 2262–2273, doi:10.1109/TMI.2013.2279938.
6. Wu, W.J.; Lin, S.W.; Moon, W.K. Combining support vector machine with genetic algorithm to classify ultrasound breast tumor images. *Comput. Med Imaging Graph.* **2012**, *36*, 627–633. doi:10.1016/j.compmedimag.2012.07.004.
7. Moon, W.K.; Lo, C.M.; Chang, J.M.; Huang, C.S.; Chen, J.H.; Chang, R.F. Quantitative Ultrasound Analysis for Classification of BI-RADS Category 3 Breast Masses. *J. Digit. Imaging* **2013**, *26*, 1091–1098, doi:10.1007/s10278-013-9593-8.

8. Shan, J.; Alam, S.K.; Garra, B.; Zhang, Y.; Ahmed, T. Computer-Aided Diagnosis for Breast Ultrasound Using Computerized BI-RADS Features and Machine Learning Methods. *Ultrasound Med. Biol.* **2016**, *42*, 980–988, doi:10.1016/j.ultrasmedbio.2015.11.016.
9. Yu, Y.; Acton, S.T. Speckle reducing anisotropic diffusion. *IEEE Trans. Image Process.* **2002**, *11*, 1260–1270, doi:10.1109/TIP.2002.804276.
10. Savage, S.L.; Lawrence, R.L.; Squires, J.R. Mapping post-disturbance forest landscape composition with Landsat satellite imagery. *For. Ecol. Manag.* **2017**, *399*, 9–23, doi:10.1016/j.foreco.2017.05.017.
11. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems—Volume 1*; Curran Associates Inc.: Dutchess County, NY, USA, 2012; pp. 1097–1105.
12. Zhang, W.; Li, R.; Deng, H.; Wang, L.; Lin, W.; Ji, S.; Shen, D. Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. *NeuroImage* **2015**, *108*, 214–224, doi:10.1016/j.neuroimage.2014.12.061.
13. Havaei, M.; Davy, A.; Warde-Farley, D.; Biard, A.; Courville, A.; Bengio, Y.; Pal, C.; Jodoin, P.M.; Larochelle, H. Brain tumor segmentation with Deep Neural Networks. *Med. Image Anal.* **2017**, *35*, 18–31, doi:10.1016/j.media.2016.05.004.
14. Yap, M.H.; Pons, G.; Martí, J.; Ganau, S.; Sentís, M.; Zwigglelaar, R.; Davison, A.K.; Martí, R. Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 1218–1226, doi:10.1109/JBHI.2017.2731873.
15. Chiang, T.; Huang, Y.; Chen, R.; Huang, C.; Chang, R. Tumor Detection in Automated Breast Ultrasound Using 3-D CNN and Prioritized Candidate Aggregation. *IEEE Trans. Med. Imaging* **2019**, *38*, 240–249, doi:10.1109/TMI.2018.2860257.
16. Acharya, U.R.; Meiburger, K.M.; Wei Koh, J.E.; Ciaccio, E.J.; Arunkumar, N.; Hoong See, M.; Mohd Taib, N.A.; Vijayanathan, A.; Rahmat, K.; Fadzli, F.; et al. A Novel Algorithm for Breast Lesion Detection Using Textons and Local Configuration Pattern Features With Ultrasound Imagery. *IEEE Access* **2019**, *7*, 22829–22842, doi:10.1109/ACCESS.2019.2898121.
17. Perona, P.; Malik, J. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* **1990**, *12*, 629–639, doi:10.1109/34.56205.
18. Marcomini, K.D.; Schiabel, H.; Carneiro, A.A.O. Quantitative evaluation of automatic methods for lesions detection in breast ultrasound images. In *Medical Imaging 2013: Computer-Aided Diagnosis*; Novak, C.L., Aylward, S., Eds.; SPIE Medical Imaging; International Society for Optics and Photonics: Orlando, FL, USA, 2013; Volume 8670, pp. 569–575.
19. Poonguzhali, S.; Ravindran, G. A complete automatic region growing method for segmentation of masses on ultrasound images. In *Proceedings of the 2006 International Conference on Biomedical and Pharmaceutical Engineering*, Singapore, 11–14 December 2006; pp. 88–92.
20. Chang, C.C.; Lin, C.J. LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol. (TIST)* **2011**, *2*, 27:1–27:27, doi:10.1145/1961189.1961199.
21. Hsu, C.W.; Lin, C.J. A comparison of methods for multiclass support vector machines. *IEEE Trans. Neural Netw.* **2002**, *13*, 415–425, doi:10.1109/72.991427.
22. van den Burg, G.J.; Groenen, P.J. GenSVM: A Generalized Multiclass Support Vector Machine. *J. Mach. Learn. Res.* **2016**, *17*, 7964–8005.
23. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32, doi:10.1023/A:1010933404324.
24. Bosch, A.; Zisserman, A.; Munoz, X. Image Classification using Random Forests and Ferns. In *Proceedings of the 2007 IEEE 11th International Conference on Computer Vision*, Rio de Janeiro, Brazil, 14–21 October 2007.
25. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324, doi:10.1109/5.726791.

