

Article

Breast Ultrasound Computer-Aided Diagnosis System Based on Mass Irregularity Features in Frequency Domain

Tahsin Nairuz ^{1,†} , Deokwoo Lee ^{2,†}  and Jong-Ha Lee ^{1,*}

¹ Department of Biomedical Engineering, Keimyung University, Daegu 42601, Republic of Korea; tahsin.bmb@nstu.edu.bd

² Department of Computer Engineering, Keimyung University, Daegu 42601, Republic of Korea; dwoolee@kmu.ac.kr

* Correspondence: segeberg@kmu.ac.kr

† These authors contributed equally to this work.

Abstract: Our study develops a computer-aided diagnosis (CAD) system for breast ultrasound by presenting an innovative frequency domain technique for extracting mass irregularity features, thereby significantly boosting tumor classification accuracy. The experimental data consists of 5252 ultrasound breast tumor images, including 2745 benign tumors and 2507 malignant tumors. A Support Vector Machine was employed to classify the tumor as either benign or malignant, and the effectiveness of the proposed features set in distinguishing malignant masses from benign ones was validated. For the constructed CAD system, the performance indices' accuracy, sensitivity, specificity, PPV, and NPV were 92.91%, 89.94%, 91.38%, 90.29%, and 91.45%, respectively, and the area index in the ROC analysis (AUC) was 0.924, demonstrating our method's superiority over traditional spatial gray level dependence (SGLD), the ratio of depth to width, the count of depressions, and orientation features. Therefore, the constructed CAD system with the proposed features will be able to provide a precise and quick distinction between benign and malignant breast tumors with minimal training time in clinical settings.

Keywords: breast cancer; computer-aided diagnosis (CAD); Breast Imaging Reporting and Data System (BI-RADS); receiver operating characteristic (ROC); tumor classification; support vector machine (SVM)



Citation: Nairuz, T.; Lee, D.; Lee, J.-H. Breast Ultrasound Computer-Aided Diagnosis System Based on Mass Irregularity Features in Frequency Domain. *Appl. Sci.* **2024**, *14*, 8003. <https://doi.org/10.3390/app14178003>

Academic Editor: Herbert Schneckeburger

Received: 11 August 2024

Revised: 29 August 2024

Accepted: 4 September 2024

Published: 7 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The hallmark of breast cancer lies in the uncontrolled proliferation and growth of abnormal cells, which eventually appear as a distinct mass within the breast tissue. Breast cancer stands as the second leading cause of cancer-related mortality after lung cancer and the most prevalent malignancy affecting women globally [1]. According to data from the World Health Organization (WHO), this tumor affects 2.1 million women annually and constitutes the predominant cause of cancer-related deaths for females between the ages of 40 and 55 [2]. In fact, this particular malignancy is responsible for 15% of all cancer-related deaths within the female population [2]. With the advancements in pharmaceuticals yielding new treatments with improved efficacy and reduced toxicity, breast cancer has evolved into a malignancy with a better prognosis. Consequently, the paradigm for the prevention and treatment of breast cancer has shifted towards managing it as a chronic condition, necessitating an extensive assortment of strategies, from screening to post-diagnostic treatment and follow-up. Therefore, early detection of breast cancer is crucial to the overarching management framework, improving treatment success rates and substantially reducing mortality rates [3,4].

For the adjunctive diagnosis of breast cancer, imaging modalities such as mammography (MG), ultrasound (US), magnetic resonance imaging (MRI), computed tomography

(CT), and positron emission tomography (PET) serve as indispensable tools, each offering distinct insights into anatomical structures, functional characteristics, and metabolic processes, thereby facilitating comprehensive clinical decision-making [5]. Nevertheless, each diagnostic modality presents inherent drawbacks. Mammography is predominantly employed in the screening of asymptomatic women with breast cancer, leveraging its capability to detect small tumors [6,7] using minimal X-ray exposure [8]. Despite its extensive adoption, mammography has some limitations, including a higher likelihood of false positives [9,10]. This can lead to a lower specificity, resulting in unnecessary biopsies and making the screening process more complicated [10]. Moreover, MRI exhibits higher sensitivity compared to the aforementioned techniques in detecting breast cancer, albeit at a higher cost and with relatively lower specificity [11,12]. Conversely, imaging modalities like PET and CT entail potential risks associated with radiation exposure [13].

In the realm of clinical practice, ultrasound emerges as the preeminent imaging modality for the staging of tumors and guidance of biopsies in instances of diagnosed or suspected primary breast cancer, largely due to its remarkable versatility, portability, and cost-effectiveness [14]. Despite its reliance on the expertise of highly trained operators, ultrasounds are increasingly esteemed as a critical adjunct in routine breast cancer screenings, attributed to its superior capability in visualizing dense breast tissue, devoid of the risks associated with ionizing radiation. The utility of ultrasound is particularly pronounced in delineating breast masses situated in technically challenging regions for other imaging methods, such as the margins of medial quadrants adjacent to the breastbone and in aberrant mammary lobes. Furthermore, ultrasound uniquely excels in the comprehensive analysis of whole-breast vascularity and the vascular patterns of lesions, alongside its precision in characterizing abnormal lymph nodes. The method's noninvasive, harmless nature, coupled with its accessibility and considerable effectiveness in diagnosing a broad spectrum of breast conditions, underpins its frequent application even in the absence of specific indications, often serving preventive purposes.

Notwithstanding its numerous advantages, the quality of ultrasound imaging has been hampered by inherent speckle noise and low contrast discernibility in different types of tissues. To ameliorate these issues, a spectrum of digital image processing techniques and advanced machine learning methodologies have been applied, thereby enhancing detection rates and specificity [15–17]. The swift development in computer application technologies has precipitated a paradigm shift towards the reliance on computer-aided diagnosis (CAD) systems for the expeditious processing of digitized imaging data. These advancements facilitate the attenuation of background noise, the enhancement of image contrast, the precise detection of regions of interest, and the differentiation of tumors from surrounding tissues, thus aiding in the distinction between benign and potentially malignant lesions [18–21]. Among the multifaceted functionalities of CAD systems, the paramount objective remains the accurate classification of tumors into benign or suspicious categories.

To encode a mass within a computational framework, sonographic features are converted into computerized sonographic descriptors. The morphological formation of a mass can be elucidated through the characterization of its shape and margin. Mass shape, which is one of the most significant image features, can be quantified using moments and Fourier descriptors [22] or by evaluating factors such as roundness, aspect-ratio, convexity, and solidity [23]. The shapes of malignant breast tumors are typically more complex than benign lesions due to their tendency to infiltrate surrounding tissues. While malignant lesions are often irregular, microlobulated, or spiculated, benign lesions tend to be oval, round, or macrolobulated. Therefore, accurately quantifying the irregularity of sonographic findings can significantly enhance the accuracy of CAD systems. Several boundary-based methods, analyzed in the spatial domain, have been used to describe mass shape, such as roundness, concaveness, and convexity [24,25]. However, these methods primarily express shape irregularity at a local level and fail to account for comprehensive shape dynamics. For example, if a shape exhibits irregular characteristics only in certain local

areas, while most other regions appear regular, boundary-based methods may incorrectly classify the shape as irregular. Consequently, there is a need to develop advanced feature extraction techniques that can effectively characterize shape irregularity across the entire lesion structure.

The Breast Imaging Reporting and Data System (BI-RADS), an initiative developed by the American College of Radiology, meticulously standardizes the lexicon and evaluative criteria for clinical findings [26]. The BI-RADS lexicon is designed to standardize the reporting of mammographic and ultrasound findings, ensuring that the resultant reports are clear, succinct, and consistent among different readers. Although the terms within the BI-RADS framework are inherently descriptive rather than quantitative, there is a necessity to “translate” these descriptors into quantifiable features that a computer-aided diagnosis (CAD) system can automatically compute. To address this, several research groups have proposed sophisticated methodologies for the quantification of BI-RADS features [27–29]. Moreover, BI-RADS sonographic characteristic features, when correlated with patient medical history, as shown by Baker et al. [30], or with patient age, as demonstrated by Lo et al. [31], can serve as the foundation for constructing a CAD system utilizing artificial Neural Networks. Additionally, Baker et al. [32] explored the reliability of CAD system predictions compared to the clinical biopsy decisions made by five radiologists, revealing that a CAD system leveraging BI-RADS evaluations significantly enhances radiologists’ interpretation of abnormalities. The primary advantage of integrating BI-RADS sonographic evaluations within CAD systems lies in their versatile applicability across diverse ultrasound platforms [25].

For the breast ultrasound CAD system, a Support Vector Machine (SVM) is extensively employed as a preferred feature-based classification algorithm due to its effectiveness in predicting and diagnosing breast cancer, achieving optimal performance by utilizing various performance metrics. Moreover, the SVM model is considered as a reliable classifier in a CAD system because they train effectively in less time and offer efficient computational performance [33]. Several studies also demonstrated the superiority of the SVM classifier over other machine learning classification algorithms such as Random Forest, Decision Trees, Neural Network, Logistic Regression, etc. [33–36]. A recent study by Kumar et al. [35] found that SVM achieved a higher accuracy of 97.66%, along with better results in other performance metrics, in comparison to other classifiers. They also marked that SVM takes less time, and the area under curve is bigger compared to other classifiers. Furthermore, in a comparative analysis, SVM classifier outperformed other algorithms such as AdaBoost, Logistic Regression, Neural Networks, and Random Forest, achieving the highest accuracy of 98.2456%, marking it as the most effective algorithm for categorizing breast cancer cases [36].

However, a paramount challenge for a CAD based classification system lies in the discovery and extraction of highly efficient and sophisticated computerized features that can proficiently differentiate between benign and malignant tumors. Moreover, employing a CAD system based on shape analysis is challenging as it is often time-consuming due to the extensive computational effort required for extracting morphological features and training the classifier. In addition, the limitation of diagnostic accuracy was often more attributable to the algorithms’ insufficient feature extraction and computational capabilities rather than the quality of the image data itself. To develop a quick and efficient CAD system, this study introduces a cutting-edge technique for extracting mass irregularity features from ultrasound images in the frequency domain, enhancing CAD-based classification by capturing subtle shape variances from a comprehensive perspective. This study explores a method for extracting multiple sonographic features and modeling border irregularities, distinguishing nuanced shape and characteristic differences between benign and malignant lesions to improve CAD classification accuracy.

2. Materials and Methods

2.1. Ultrasound Image Acquisition and Data Collection

A comprehensive collection of 5252 breast ultrasound images was obtained from public datasets which were recorded in Samsung Medical Center, Seoul, Republic of Korea, consecutively from 2006 to 2012. A detailed description of the dataset is represented in Table 1. Among the image dataset, 2745 comprised benign breast tumors and 2507 malignant breast tumors. The histopathological characteristics of all breast lesions were confirmed through biopsy. The study cohort comprises only those with Asian ethnicity. The patients with benign tumors had a mean age of 45 years, with an age range extending from 11 to 81 years. Conversely, the cohort with malignant tumors had a mean age of 49 years, spanning ages from 24 to 86 years. All experimental protocols received approval from Samsung Medical Center, Seoul, Republic of Korea. Informed consent was meticulously obtained from all patients, ensuring their agreement to use their information for research purposes while maintaining privacy. These images were acquired using a Philips ATL iU22 ultrasound system (Philips, Bothell, WA, USA), operating under the authorization of the institutional review board of Samsung Medical Center. The ultrasound system employed a linear probe with a frequency range of 5 to 12 MHz and a size of 6 cm. The B-mode ultrasound images boasted a resolution of 1024 by 768 pixels, yielding a spatial detail of 0.23 mm per pixel.

Table 1. Dataset Description.

Attribute	Details
Total Number of Images	5252
Time Frame	2006–2012
Number of Benign Cases	2745
Number of Malignant Cases	2507
Mean Age of Patients (Benign Tumors)	45 years
Age Range of Patients (Benign Tumors)	11–81 years
Mean Age of Patients (Malignant Tumors)	49 years
Age Range of Patients (Malignant Tumors)	24–86 years
Ultrasound Imaging Device	Philips ATL iU22
Image Resolution	1024 × 768 pixels
Spatial Resolution	0.23 mm per pixel
Frequency Range	5–12 MHz

2.2. Fundamental Operational Framework of Breast Ultrasound CAD

The breast ultrasound CAD system is divided into two distinct modules: CADe (computer-aided detection), which delineates suspected lesion areas, and CADx (computer-aided diagnosis), which assesses the malignancy of these lesions (Figure 1). Initially, CADe is employed to identify and mark potential lesion regions. Subsequently, CADx takes over, characterizing and diagnosing the lesions based on BI-RADS criteria. The BI-RADS categories utilized in CADx serve a dual purpose: aiding the physician's diagnostic process and calculating lesion malignancy probabilities. Diagnosing a lesion's malignancy necessitates a pre-established automatic diagnostic model, which is developed using BI-RADS annotations derived from an extensive dataset of breast ultrasound images stored in the PACS (Picture Archiving and Communication System). This process begins with CADe marking and presenting suspected lesion areas within the ultrasound images, followed by the calculation and display of results in accordance with the BI-RADS lexicon.

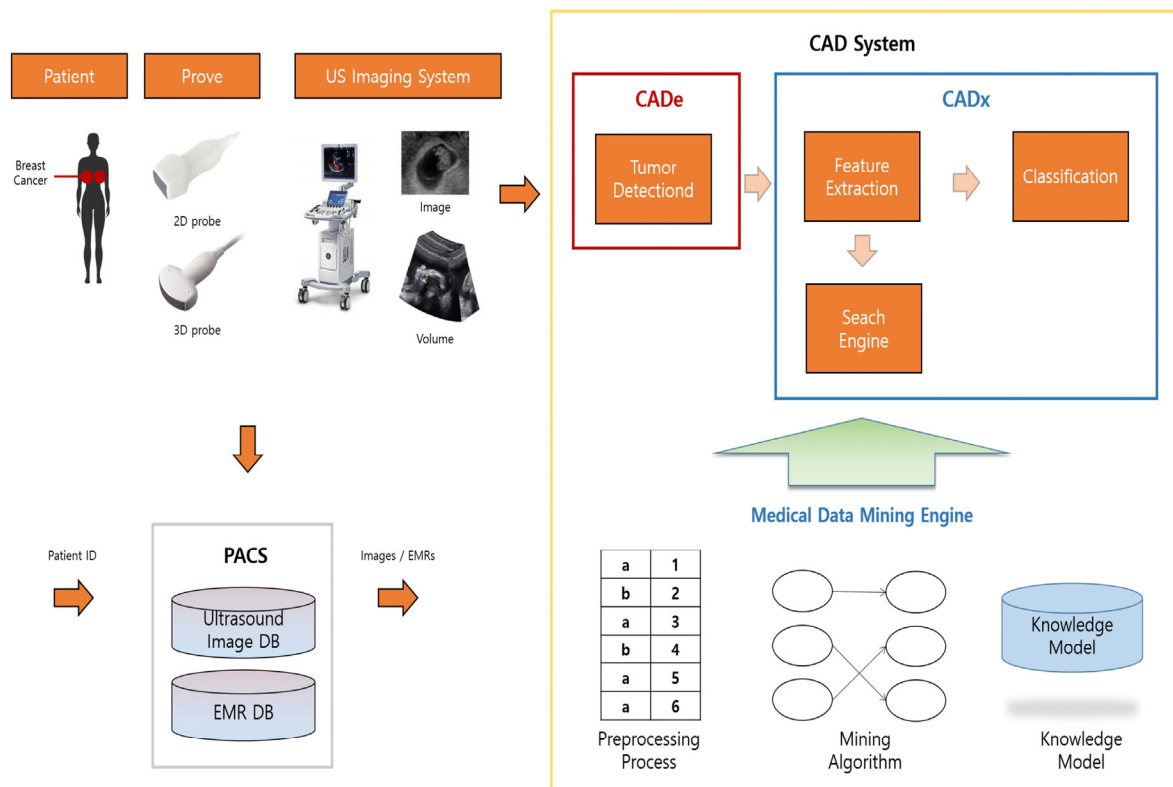


Figure 1. The conceptual architecture of the breast ultrasound CAD framework.

2.3. BI-RADS Categories

In the fifth edition of BI-RADS Ultrasound [37], the descriptive terms for characterizing the main sonographic attributes of a mass are standardized and classified into five categories: shape, orientation, margin, echo pattern, and posterior acoustic features.

Shape: In the BI-RADS framework, the morphological characterization of a lesion is crucial for assessing breast cancer. The classification of lesion shape is delineated into three specific categories: oval, which describes lesions with an elliptical form and 2–3 curvatures; round, indicating a circular configuration; and irregular, which indicates shapes that do not conform to either oval or round (Figure 2). Typically, benign tumors, which are non-proliferative and lack metastatic potential, exhibit smooth, well-defined shapes, often manifesting as oval or round. Conversely, malignant tumors display irregular shapes, a reflection of their invasive growth and metastatic properties.

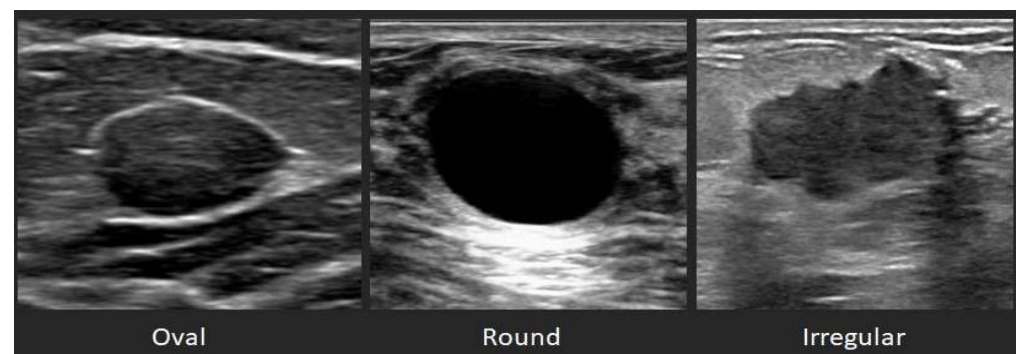


Figure 2. Examples of three mass shapes in ultrasound images.

Orientation: Orientation pertains to the alignment of the tumor's long axis. When this long axis runs parallel to the skin line, the orientation is termed parallel, or "wider

than tall”. Conversely, if the long axis is perpendicular to the skin line, the orientation is described as not parallel, or “taller than wide” (Figure 3). The “taller than wide” orientation is particularly concerning, as it suggests malignancy due to the tumor’s reduced compressibility and its propensity to infiltrate across tissue planes.

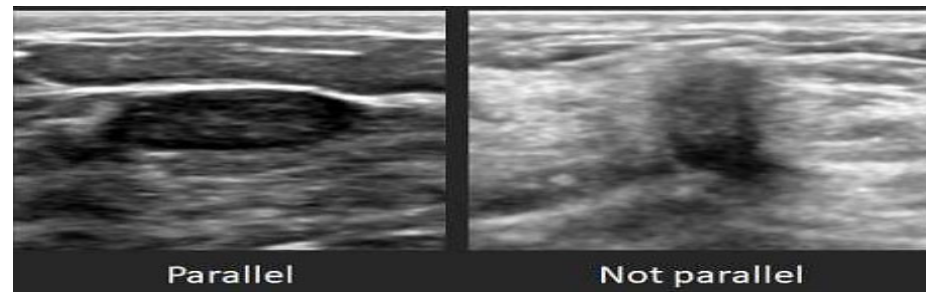


Figure 3. Example of mass orientation: parallel and not parallel.

Margin: The margin category refers to the delineation between a lesion and the surrounding tissue. When this boundary is distinct and sharply defined, it is labeled circumscribed. On the other hand, if the boundary is unclear, it falls under the not circumscribed category. Not-circumscribed margins are further divided into specific subcategories: indistinct, where the boundary is blurred and poorly defined; angular, where edges are sharp and form acute angles; microlobulated, featuring short, radiating protrusions along the lesion’s boundary; and spiculated, characterized by needle-like projections (Figure 4). The presence of one or more of these subcategories results in a not-circumscribed classification. Benign tumors, typically enclosed by a fibrous capsule, exhibit circumscribed margins due to their clear separation from surrounding tissues. In contrast, malignant tumors, which lack such encapsulation, frequently present with not circumscribed margins, reflecting their invasive nature.

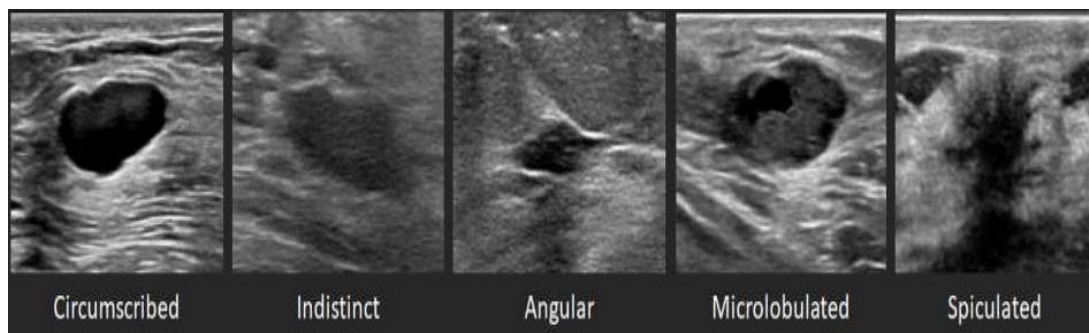


Figure 4. Examples of mass margin categories in ultrasound images.

Echo pattern: The internal echo pattern of a mass encompasses several distinct classifications: anechoic, hyperechoic, complex cystic and solid, hypoechoic, isoechoic (equal), and heterogeneous (Figure 5). Echogenicity is determined in relation to subcutaneous fat. An anechoic mass, devoid of internal echoes, presents as entirely black on ultrasound. Hyperechoic masses exhibit increased echogenicity, isoechoic masses show echogenicity comparable to subcutaneous fat, and hypoechoic masses display reduced echogenicity. Complex cystic and solid masses contain both anechoic cystic regions and echogenic solid elements, such as intracystic papillomas. A heterogeneous echo pattern is characterized by a combination of varying echogenicities within a single solid mass, reflecting a diverse internal texture.

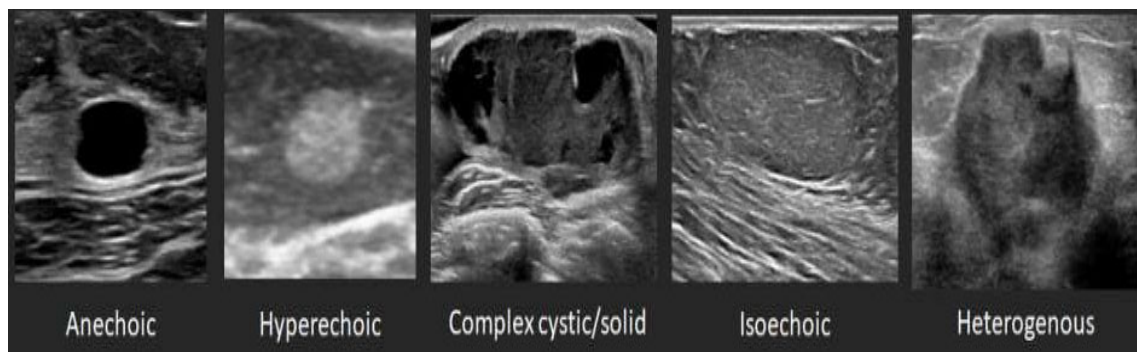


Figure 5. Examples of echo patterns in ultrasound images.

Posterior acoustic features: The posterior acoustic feature of a lesion is categorized into several distinct patterns: enhancement, which refers to an increase in echogenicity behind the tumor; shadowing, characterized by a reduction in echogenicity (excluding edge artifacts such as border shadow); absence of posterior acoustic features, where neither enhancement nor shadowing is observed; and a combined pattern, where there is a presence of both attenuation and either enhancement or shadowing at the posterior aspect (Figure 6).



Figure 6. Examples of posterior features in ultrasound images.

In this study, feature extractors were created for BI-RADS categories to calculate feature points from images according to the purpose of each category. To generate an automatic lesion diagnosis model, feature points for each BI-RADS category were calculated using designed feature extractors from a large volume of breast ultrasound images in PACS. Then, pattern recognition algorithms and data mining techniques were employed to create a breast cancer auxiliary diagnosis model. For the pattern recognition algorithm, we utilized a Support Vector Machine (SVM) classifier due to its ability to handle high-dimensional feature spaces and its effectiveness in binary classification—critical for accurately distinguishing between benign and malignant breast lesions. The data mining process began with the extraction of key features from a large set of breast ultrasound images stored in the PACS system. For each BI-RADS category, we developed specific feature extractors designed to capture relevant image characteristics that are indicative of different types of lesions. These extracted features were then standardized and processed to reduce noise and improve their utility for classification tasks. The refined feature set served as input for the SVM classifier, which was trained to detect patterns that correspond to either benign or malignant lesions based on the features of each BI-RADS category. By integrating these methodologies, this developed an auxiliary diagnosis model for breast cancer which locates the lesion and automatically determines its benignity or malignancy when new breast ultrasound images are input into the CAD system.

2.4. Lesion Detection and Contour Extraction

2.4.1. Detection of Lesion Areas Using Morphological Information of the Lesion

Breast ultrasound imaging is intrinsically characterized by a pervasive presence of speckle noise and can be systematically delineated into distinct anatomical features, including the skin, subcutaneous fat, mammary glandular tissue, retromammary fat, and the pectoralis muscle. In these images, the fat layers and lesions manifest as dark regions, whereas the breast parenchyma exhibits a brighter appearance. Consequently, lesions are typically discernible by their darker contrast relative to the surrounding tissue and are often circular or elliptical. Notably, these lesions predominantly reside within the mammary glandular tissue. Therefore, by confining the detection process to the mammary glandular tissue area, as opposed to scanning the entire image, the efficacy of lesion detection can be potentially enhanced. This study delves into the development of an algorithm tailored for lesion detection within the mammary glandular tissue, leveraging both the morphological characteristics of the lesions and the structural intricacies of the breast (Figure 7).

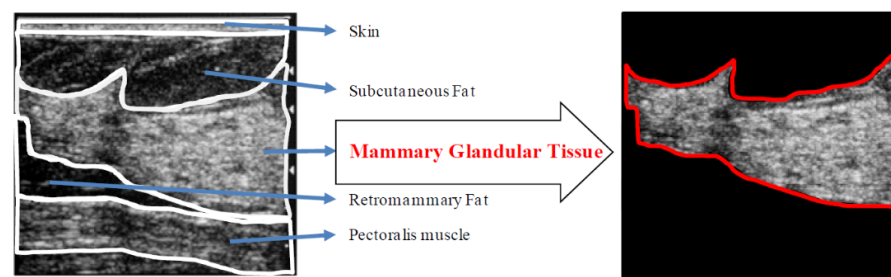


Figure 7. Examples of morphological information from breast ultrasound images and segmented boundaries of breast lesion.

2.4.2. Lesion Contour Extraction Using the Canny Algorithm

After detecting the lesion area, it is necessary to extract the lesion's contour. Generally, in breast ultrasound images, the area of a lesion has darker characteristics compared to its surroundings, leading to a sharp change in pixel values within the image. In this study, we extract the lesion's contour using the gradient information of pixel values. First, the image is convolved with a Gaussian mask. The Gaussian mask is a type of blurring technique that smoothens the image by removing its fine details. Next, the gradient magnitude is calculated for each pixel, and the direction of the gradient at each pixel is determined. Following this, the second derivative is calculated along the gradient direction, and the locations with a value of zero are identified as the contour of the lesion. By extracting the contour of the lesion at each point and then connecting these points, the final contour of the lesion is extracted.

2.5. Tumor Malignancy Determination Using BI-RADS Category Information

This study explores an algorithm to analyze lesions based on their extracted contours. We assess the malignancy of lesions in breast ultrasound images by utilizing data from five BI-RADS categories. The BI-RADS categories of interest encompass shape (regular/irregular), orientation (parallel/not parallel), margin (circumscribed/not circumscribed), echo pattern (anechoic/hyperechoic), and posterior features (no posterior/shadowing). These categorical attributes, crucial for the assessment of masses, play a pivotal role in lesion diagnosis. To delineate each BI-RADS category, we leverage both morphological characteristics and textural features of the lesion. Shape irregularity is particularly critical in determining its malignancy and is quantified using a Fourier transform-based shape description algorithm. This approach converts the spatial coordinates of the tumor contour into frequency domain features, which are then analyzed to assess irregularity.

2.6. Mass Irregularity Feature Extraction

Initially, control points are selected from the segmented tumor boundary, and shape attributes are derived using a Fourier descriptor based on the centroid shape context function (CSCF). The CSCF analysis commences with the selection of N points along the shape's contour. This process involves generating a set of vectors from a designated starting point to each subsequent point along the contour, thereby encapsulating the overall shape configuration relative to a reference point. The CSCF method captures the shape's entirety in relation to the mass's centroid through a log-polar representation, and it is depicted via a shape histogram.

In the log-polar histogram, bins are uniformly spaced to enhance sensitivity to the positioning of adjacent points rather than distant ones. For instance, a log-polar histogram, such as illustrated in Figure 8a, might include 5 bins for the radial direction (r) and 12 bins for the angular direction (θ). By centering the histogram at the mass's centroid, each bin accumulates the count of contour points that fall within its spatial domain. The histogram data from each bin is organized into a two-dimensional (2-D) histogram. When centered on the centroid, this 2-D histogram can be reorganized into a one-dimensional (1-D) histogram (Figure 8b,c). This conversion is achieved by sorting the data based on decreasing polar distance.

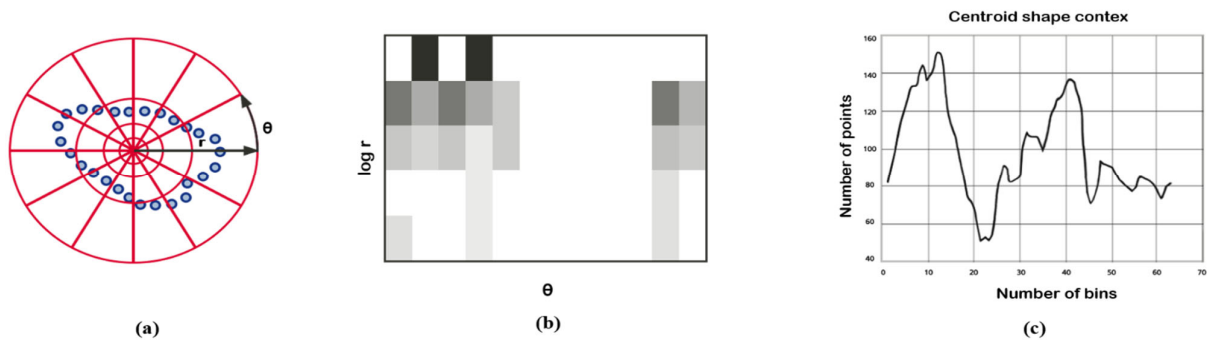


Figure 8. (a) Log-polar diagram, (b) two-dimensional histogram of boundary points included per bin, (c) histogram converted into one-dimensional.

The Fourier descriptor (FD) is subsequently employed to analyze and represent the closed plane curves. This descriptor is derived through the Fourier transform applied to the shape's boundary coordinates, resulting in normalized coefficients from this transformation. This FD is acquired by conducting a Fourier transform on a shape signature, which originates from the coordinates along the shape's boundary. The Fourier descriptors refer to the normalized coefficients that result from this Fourier transform process. The discrete Fourier transform of the signature function is given by

$$a_n = \frac{1}{N} \sum_{t=0}^{N-1} r(t) e^{-j2\pi \frac{nt}{N}}, \quad n = 0, 1, \dots, N-1 \quad (1)$$

Given that the centroid shape context function (CSCF) is unaffected by changes in position or orientation, adjustments to the Fourier coefficients are necessary to ensure they are independent of the scale and the initial point of the shape descriptors. The mathematical expression linking the Fourier coefficients of the shape in its original form to its version altered by scaling, and modifications in the starting point are maintained as

$$a_n = \exp(jn\varphi) \cdot s \cdot a_n^0 \quad (2)$$

where φ denotes the angle adjustments resulting from altering the starting point, and s represents the scaling factor. Following this, the normalized Fourier coefficients b_n for the shape after transformation are derived by

$$b_n = \frac{a_n}{a_1} = \frac{a_n^0}{a_1^0} \exp [j(n-1)\varphi] = b_n^0 \exp [j(n-1)\varphi] \quad (3)$$

where b_n^0 represents the normalized Fourier coefficients pertaining to the shape's original form. Based on Equation (3), it becomes evident that disregarding the phase component allows the magnitudes b_n and b_n^0 to remain identical. This implies that b_n maintains its consistency across changes in translation, rotation, scale, and the initial starting point. The magnitude values b_n are utilized as the distinguishing features for classifying breast tumors in our study.

2.7. Correlation-Based Feature Selection Approach

In this study, we extracted 94 attributes from each region of interest (ROI) in a dataset of 5252 breast ultrasound images. These attributes were designed to capture various shape, texture, and intensity features relevant to distinguishing between benign and malignant lesions. To refine the feature set, a correlation-based feature selection (CFS) approach was applied in this study. This method is designed to select subsets of features that are highly correlated with the class label (indicating benign or malignant lesions) while maintaining low inter-correlation among the features themselves.

During the feature selection process, we used a best-first search strategy in conjunction with the CFS approach to explore different subsets of features. This strategy repetitively added features that maximized a defined evaluation criterion based on the correlation between features and the class labels while minimizing redundancy. For example, from the initial features, several key features from different groups were selected for their high relevance and low correlation with other features. For instance, spatial gray-level dependence matrix (SGLD) features were extensively evaluated, given their ability to capture texture details crucial for distinguishing lesions. Similarly, Fourier features with shape context and centroid distance were carefully assessed for their shape descriptors, which provide critical insights into the lesion's contour and structure [38].

Through this selection process, we refined the original set of 94 features down to the most relevant 43 features which are listed in Table 2. These selected features were those that demonstrated the highest predictive power for classifying the lesions as benign or malignant.

Table 2. List of sonographic features considered in this study.

Feature No.	Feature Name
1~8	Spatial gray-level dependence matrix (SGLD)
9~16	Fourier with shape context
17~20	Fourier with centroid distance (magnitude)
21~24	Fourier with centroid distance (phase)
25	Intensity in the mass area
26	Gradient magnitude in the mass area
27	Orientation
28	Depth–width ratio
29~30	Distance between mass shape and best fit ellipse
31	The average gray changes between tissue area and mass area
32	The average gray changes between posterior and mass area
33~34	The histogram changes between tissue and mass
35	Comparison of the gray value of left, post, and right under lesion
36	The number of lobulate areas
37	The number of protuberances
38	The number of depressions
39	Lobulation index
42~43	Elliptic-normalized circumference

2.8. Support Vector Machine (SVM) Classifier

The support vector machine (SVM) technique is extensively applied for both classification and regression problems. In binary classification scenarios, such as the one addressed in this study, the goal is to determine an ideal separating hyperplane. Considering a binary classification with training data $\{x_i, y_i\}$, $i = 1, 2, \dots, N$, where $x_i \in R^d$ are the feature vectors in a d -dimensional feature space and $y_i \in \{-1, 1\}$ are the corresponding binary labels, the formula for the decision function that defines the separating hyperplane in the case of linearly separable classes is described as

$$f(x) = \omega \cdot \Phi(x) + b$$

where $\omega \in R^d$ represents the hyperplane's normal vector, and $b \in R$ is the offset. In this context, $\Phi(x)$ is the kernel function that maps the original feature space into a higher-dimensional space. The Radial Basis Function (RBF) kernel, a common choice for SVMs, allows the SVM to create non-linear decision boundaries by mapping input features into a higher-dimensional space. The SVM aims to position a hyperplane within this transformed feature space that maximizes the margin between the two categories and minimizes the classification error. The samples closest to the hyperplane are termed support vectors. This study utilizes an SVM classifier equipped with the RBF kernel for the classification tasks, marking its ability to handle non-linear relationships between features effectively.

3. Results

In our experiment, to evaluate the efficacy of the derived shape features, classification tests were carried out on the 5252 breast ultrasound images, from which 94 attributes were extracted for each image's region of interest (ROI). Subsequently, feature selection was performed through an exploration of various subsets of features, assessing each to identify the most effective combination. This process resulted in a refined set of 43 features, selected from the original pool using a correlation-based feature selection approach coupled with a best-first search strategy.

In this study, the computational experiments were conducted on a workstation equipped with a 2.93 GHz Intel Xeon CPU, 3 GB of RAM (Intel, Santa Clara, CA, USA), and running the Windows 7 OS. The k-fold cross-validation technique [39] was utilized on the entire database to determine the classification accuracy of CAD system. In this study, with k set to 10, the adopted cases were randomly partitioned into 10 sets according to the pathological result. For each round of experiments, nine sets are regarded as the training set and the remaining one as the test set, i.e., 80% of the cases were used as training sets, and 20% were testing sets. In each round of experiments, 15% of the training sets were used as the validation dataset. The model's performance was observed based on the validation dataset to adjust the model's learning rate and determine when to stop training to avoiding overfitting the model. Finally, the performance of the CAD system was evaluated by averaging the ten experiments.

While pathological results served as the gold standard, the performance of the experimental result could be quantified using five key metrics: accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) (Table 3).

Table 3. List of Five Designated Performance Metrics.

Performance Indices	Formula
Accuracy	$(TP + TN) / (TP + TN + FP + FN)$
Sensitivity	$TP / (TP + FN)$
Specificity	$TN / (TN + FP)$
Positive Predictive Value (PPV)	$TP / (TP + FP)$
Negative Predictive Value (NPV)	$TN / (TN + FN)$

TP, true positive; TN, true negative; FP, false positive, and FN, false negative.

In our study, the performance of the proposed features compared with other shape feature sets: the spatial gray level dependence (SGLD) feature (F1), depth–width ratio feature (F2), number of depressions feature (F3), and the orientation feature (F4), across five metrics to assess the effectiveness of the newly developed features (Figure 9).

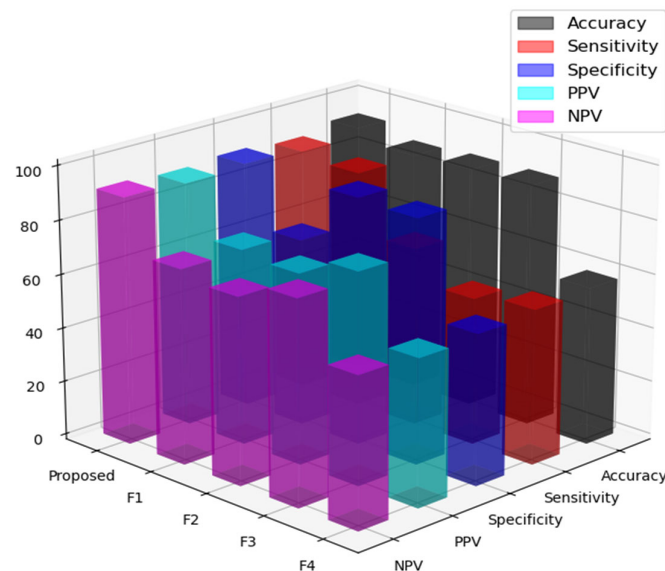


Figure 9. Classification performance of breast tumor using five objective indices.

The outcomes summarized in Table 4 indicate that the proposed features stand out with the highest accuracy at 92.91% compared to other feature sets, which follow with an accuracy of 88.42% (F1), 89.32% (F2), 89.52% (F3), and 58.29% (F4), respectively. In terms of sensitivity and specificity, the proposed features outperform all other feature sets with a rate of 89.94% and 91.38%, respectively, indicating the ability to correctly identify positive and negative cases. Moreover, the proposed features set shows strong results in PPV (90.29%), and NPV (91.45%). Hence, the findings reveal that the introduced features enhance the classification of breast tumors with greater precision across the five designated performance metrics.

Table 4. Comparative Performance Analysis of Various Feature Sets.

Feature Set	Number of Features	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Proposed Features	43	92.91	89.94	91.38	90.29	91.45
SGLD (F1)	8	88.42	87.89	69.24	72.33	71.95
Depth–Width Ratio (F2)	1	89.32	65.97	91.28	70.39	68.87
Number of Depressions (F3)	1	89.52	54.28	90.24	77.87	75.27
Orientation (F4)	1	58.29	57.22	55.56	54.17	55.56

The correlation between sensitivity and specificity is depicted through the receiver operating characteristic (ROC) curve, which plots the true positive (TP) rate versus the false positive (FP) rate. The ROC curve offers a sophisticated tool for examining a diagnostic method's classification precision. The area under the curve (AUC) defines the ROC curve's form, making the AUC a valuable indicator of testing precision (Figure 10). An AUC value of 1 signifies an ideal test, while an AUC of 0.5 indicates a test without diagnostic value.

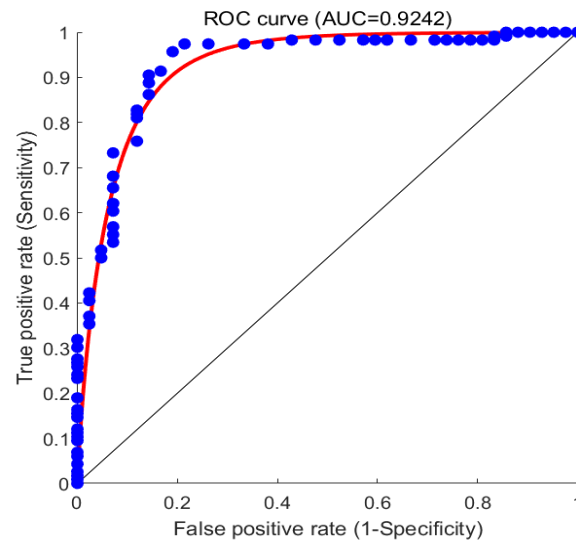


Figure 10. ROC curve of the CAD system using the proposed features and an area index of 0.9242.

An experimental outcome indicating a larger AUC suggests a superior method compared to those with lesser areas. Figure 11 illustrates the classification efficacy of our suggested features against others, as measured by AUC. The result demonstrates that the AUC for our suggested features stands at 92.4% and surpasses that of other features, such as F_1 , F_2 , F_3 , and F_4 , by 6.8%, 10.1%, 10.5%, and 34.6%, respectively. Hence, our proposed approach exhibits the highest capability for discrimination.

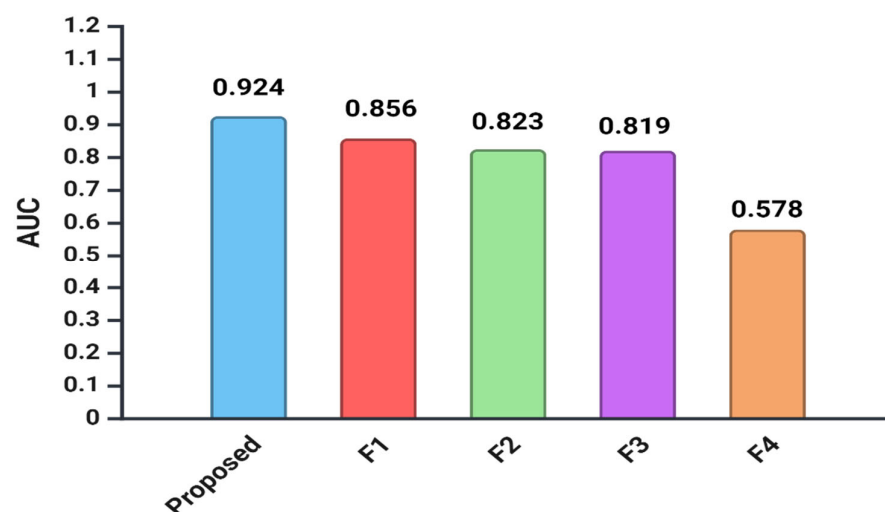


Figure 11. Evaluating the AUC from classification outcomes across various feature sets.

4. Discussion

The rapid evolution of ultrasound technology has positioned it as a key tool for diagnosing breast tumors. However, the diagnostic accuracy of ultrasounds remains disputed due to the overlapping features of benign and malignant lesions in ultrasonographic images, leading to subjective interpretations influenced by the operator's skill. To address this, advanced CAD systems have been developed, offering critical support in interpreting sonographic images of tumors. These CAD systems enhance diagnostic accuracy, especially for less experienced physicians, by analyzing intricate and multifaceted image features, thus mitigating subjectivity and improving differentiation between benign and malignant tumors.

In our study, we investigate a method for extracting features based on mass irregularity within the frequency domain, enhancing the accuracy of mass classification in breast tumor

diagnosis via computer-aided B-mode ultrasound imaging. To assess the effectiveness of these newly developed features, we utilized a dataset of 5252 ultrasound images, which included 2507 cases of malignant tumors and 2745 cases of benign tumors. Furthermore, we explored various feature types, evaluating and comparing their efficacy in tumor classification. Unlike the approach by Chang et al. [23] we utilized morphological features, significantly reducing complexity and extensive computation. Simultaneously, by incorporating texture features, our proposed system maintained high accuracy in diagnosing breast tumors.

This experiment indicates that the proposed features achieved the best performance, with an AUC of 92.4%, accuracy of 92.91%, sensitivity of 89.94%, and specificity of 91.38%, yielding improved classification outcomes over other spatial domain boundary-focused features. Therefore, this result signifies a high likelihood of accurately identifying malignant tumors. Additionally, high positive predictive value (90.29%) and negative predictive value (91.45%) suggest a notable decrease in the frequency of unnecessary biopsies for benign lesions. This enhancement not only boosts diagnostic confidence but also provides a valuable second opinion, mitigating the risk of misdiagnosis. Consequently, this approach can substantially lower the overall costs associated with breast cancer diagnosis in practical clinical settings.

Future research endeavors could encompass the integration of deep learning, a rapidly advancing and dynamic field that offers a sophisticated framework for the automated extraction of features via a self-learning network. Since high-quality features are fundamental to enhancing classification accuracy, the application of deep learning represents a potentially transformative approach for deriving robust and discriminative features essential for the automated classification of breast tumors. With the advancements in image processing technology, recent studies also demonstrate the effectiveness of deep learning algorithms in medical image analysis. For instance, the development of the CSwin-PNet, which integrates CNNs with Swin Transformers, has shown significant improvements in breast lesion segmentation in ultrasound images [40]. Another study introduced a magnified adaptive-feature pyramid network designed for the automatic detection of microaneurysms, demonstrating its potential to enhance image-based diagnosis [41]. Additionally, an adaptive multi-scale feature pyramid network, known as AMFP-net, has been proposed for the accurate diagnosis of pneumoconiosis from chest X-ray images [42]. These advancements underscore the potential of deep learning approaches to improve diagnostic accuracy and efficiency in breast cancer detection.

5. Conclusions

Our study advances computer-aided ultrasound diagnosis for breast cancer by introducing a novel frequency domain approach for extracting mass irregularity features, markedly enhancing tumor classification accuracy. This approach's potential to significantly reduce diagnostic variability and improve accuracy highlights its importance. However, the reliance on manual segmentation underscores the need for further automation to reduce subjectivity. Future research should focus on integrating these findings into clinical practice, assessing their impact on diagnostic workflows, and exploring the technique's applicability to other types of tumors. Our findings set a new direction for ultrasound-based diagnostics, emphasizing the importance of advanced image analysis techniques in improving breast cancer detection and treatment strategies.

Author Contributions: Conceptualization, J.-H.L., D.L. and T.N.; methodology, J.-H.L.; software, J.-H.L.; validation, J.-H.L. and T.N.; formal analysis, D.L.; investigation, J.-H.L. and D.L.; resources, J.-H.L. and T.N.; data curation, J.-H.L., D.L. and T.N.; writing—original draft preparation, J.-H.L. and T.N.; writing—review and editing, J.-H.L. and T.N.; visualization, J.-H.L., D.L. and T.N.; supervision, J.-H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Bisa Research Grant of Keimyung University in 2023 (No. 20230185).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article.

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

References

1. American Cancer Society. *Cancer Facts & Figures 2024*; American Cancer Society: Atlanta, GA, USA, 2024.
2. Huang, Q.; Ding, H.; Effatparvar, M. Breast Cancer Diagnosis Based on Hybrid SqueezeNet and Improved Chef-Based Optimizer. *Expert Syst. Appl.* **2024**, *237*, 121470. [\[CrossRef\]](#)
3. Marmot, M.G.; Altman, D.G.; Cameron, D.A.; Dewar, J.A.; Thompson, S.G.; Wilcox, M. The Benefits and Harms of Breast Cancer Screening: An Independent Review. *Br. J. Cancer* **2013**, *108*, 2205–2240. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Sun, L.; Legood, R.; Sadique, Z.; dos-Santos-Silva, I.; Yang, L. Cost-Effectiveness of Risk-Based Breast Cancer Screening Programme, China. *Bull. World Health Organ.* **2018**, *96*, 568. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Pomerantz, B.J. Imaging and Interventional Radiology for Cancer Management. *Surg. Clin. N. Am.* **2020**, *100*, 499–506. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Bevers, T.B.; Helvie, M.; Bonaccio, E.; Calhoun, K.E.; Daly, M.B.; Farrar, W.B.; Garber, J.E.; Gray, R.; Greenberg, C.C.; Greenup, R.; et al. Breast Cancer Screening and Diagnosis, Version 3.2018, NCCN Clinical Practice Guidelines in Oncology. *J. Natl. Compr. Canc. Netw.* **2018**, *16*, 1362–1389. [\[CrossRef\]](#)
7. Coleman, C. Early Detection and Screening for Breast Cancer. *Semin. Oncol. Nurs.* **2017**, *33*, 141–155. [\[CrossRef\]](#)
8. Maitra, I.K.; Nag, S.; Bandyopadhyay, S.K. Technique for Preprocessing of Digital Mammogram. *Comput. Methods Programs Biomed.* **2012**, *107*, 175–188. [\[CrossRef\]](#)
9. Pace, L.E. False-Positive Results of Mammography Screening in the Era of Digital Breast Tomosynthesis. *JAMA Netw. Open* **2022**, *5*, e222445. [\[CrossRef\]](#)
10. Guo, Z.; Xie, J.; Wan, Y.; Zhang, M.; Qiao, L.; Yu, J.; Chen, S.; Li, B.; Yao, Y. A Review of the Current State of the Computer-Aided Diagnosis (CAD) Systems for Breast Cancer Diagnosis. *Open Life Sci.* **2022**, *17*, 1600–1611. [\[CrossRef\]](#)
11. Ding, W.; Fan, Z.; Xu, Y.; Wei, C.; Li, Z.; Lin, Y.; Ruan, G. Magnetic Resonance Imaging in Screening Women at High Risk of Breast Cancer: A Meta-Analysis. *Medicine* **2023**, *102*, e33146. [\[CrossRef\]](#)
12. Chiarelli, A.M.; Blackmore, K.M.; Muradali, D.; Done, S.J.; Majpruz, V.; Weerasinghe, A.; Mirea, L.; Eisen, A.; Rabeneck, L.; Warner, E. Performance Measures of Magnetic Resonance Imaging Plus Mammography in the High Risk Ontario Breast Screening Program. *J. Natl. Cancer Inst.* **2020**, *112*, 136–144. [\[CrossRef\]](#)
13. Armanious, K.; Jiang, C.; Fischer, M.; Küstner, T.; Hepp, T.; Nikolaou, K.; Gatidis, S.; Yang, B. MedGAN: Medical Image Translation Using GANs. *Comput. Med. Imaging Graph.* **2020**, *79*, 101684. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Sood, R.; Rositch, A.F.; Shakoob, D.; Ambinder, E.; Pool, K.-L.; Pollack, E.; Mollura, D.J.; Mullen, L.A.; Harvey, S.C. Ultrasound for Breast Cancer Detection Globally: A Systematic Review and Meta-Analysis. *J. Glob. Oncol.* **2019**, *5*, 1–17. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Moinuddin, M.; Khan, S.; Alsaggaf, A.U.; Abdulaal, M.J.; Al-Saggaf, U.M.; Ye, J.C. Medical Ultrasound Image Speckle Reduction and Resolution Enhancement Using Texture Compensated Multi-Resolution Convolution Neural Network. *Front. Physiol.* **2022**, *13*, 961571. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Zhang, L.; Zhang, J. Ultrasound Image Denoising Using Generative Adversarial Networks with Residual Dense Connectivity and Weighted Joint Loss. *PeerJ Comput. Sci.* **2022**, *8*, e873. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Cammarasana, S.; Nicolardi, P.; Patanè, G. Real-Time Denoising of Ultrasound Images Based on Deep Learning. *Med. Biol. Eng. Comput.* **2022**, *60*, 2229–2244. [\[CrossRef\]](#)
18. Cao, Z.; Duan, L.; Yang, G.; Yue, T.; Chen, Q. An Experimental Study on Breast Lesion Detection and Classification from Ultrasound Images Using Deep Learning Architectures. *BMC Med. Imaging* **2019**, *19*, 51. [\[CrossRef\]](#)
19. AlZoubi, A.; Lu, F.; Zhu, Y.; Ying, T.; Ahmed, M.; Du, H. Classification of Breast Lesions in Ultrasound Images Using Deep Convolutional Neural Networks: Transfer Learning Versus Automatic Architecture Design. *Med. Biol. Eng. Comput.* **2024**, *62*, 135–149. [\[CrossRef\]](#)
20. Gu, Y.; Xu, W.; Lin, B.; An, X.; Tian, J.; Ran, H.; Ren, W.; Chang, C.; Yuan, J.; Kang, C.; et al. Deep Learning Based on Ultrasound Images Assists Breast Lesion Diagnosis in China: A Multicenter Diagnostic Study. *Insights Imaging* **2022**, *13*, 124. [\[CrossRef\]](#)
21. Sirjani, N.; Oghli, M.G.; Tarzamani, M.K.; Gity, M.; Shabanzadeh, A.; Ghaderi, P.; Shiri, I.; Akhavan, A.; Faraji, M.; Taghipour, M. A Novel Deep Learning Model for Breast Lesion Classification Using Ultrasound Images: A Multicenter Data Evaluation. *Phys. Med.* **2023**, *107*, 102560. [\[CrossRef\]](#)
22. Liang, S.; Rangayyan, R.M.; Desautels, J.E.L. Application of Shape Analysis to Mammographic Calcifications. *IEEE Trans. Med. Imaging* **1994**, *13*, 263–274. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Chang, R.F.; Wu, W.J.; Moon, W.K.; Chen, D.-R. Automatic Ultrasound Segmentation and Morphology Based Diagnosis of Solid Breast Tumors. *Breast Cancer Res. Treat.* **2005**, *89*, 179–185. [\[CrossRef\]](#)

24. Chen, C.M.; Chou, Y.H.; Han, K.C.; Hung, G.S.; Tiu, C.M.; Chiou, H.J.; Chiou, S.Y. Breast Lesions on Sonograms: Computer-Aided Diagnosis with Nearly Setting-Independent Features and Artificial Neural Networks. *Radiology* **2003**, *226*, 504–514. [[CrossRef](#)] [[PubMed](#)]
25. Shen, W.C.; Chang, R.F.; Moon, W.K.; Chou, Y.H.; Huang, C.S. Breast Ultrasound Computer-Aided Diagnosis Using BI-RADS Features. *Acad. Radiol.* **2007**, *14*, 928–939. [[CrossRef](#)] [[PubMed](#)]
26. American College of Radiology. *Breast Imaging Reporting and Data System*, 4th ed.; American College of Radiology: Reston, VA, USA, 2003.
27. 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. [[CrossRef](#)]
28. Fleury, E.; Marcomini, K. Performance of Machine Learning Software to Classify Breast Lesions Using BI-RADS Radiomic Features on Ultrasound Images. *Eur. Radiol. Exp.* **2019**, *3*, 34. [[CrossRef](#)] [[PubMed](#)]
29. Chang, Y.W.; Chen, Y.R.; Ko, C.C.; Lin, W.Y.; Lin, K.P. A Novel Computer-Aided Diagnosis System for Breast Ultrasound Images Based on BI-RADS Categories. *Appl. Sci.* **2020**, *10*, 1830. [[CrossRef](#)]
30. Baker, J.A.; Kornguth, P.J.; Lo, J.Y.; Williford, M.E.; Floyd, C.E., Jr. Breast Cancer: Prediction with Artificial Neural Network Based on BI-RADS Standardized Lexicon. *Radiology* **1995**, *196*, 817–822. [[CrossRef](#)]
31. Lo, J.Y.; Baker, J.A.; Kornguth, P.J.; Iglehart, J.D.; Floyd, C.E., Jr. Predicting breast cancer invasion with artificial neural networks on the basis of mammographic features. *Radiology* **1997**, *203*, 159–163. [[CrossRef](#)]
32. Baker, J.A.; Kornguth, P.J.; Lo, J.Y.; Floyd, C.E., Jr. Artificial neural network: Improving the quality of breast biopsy recommendations. *Radiology* **1996**, *198*, 131–135. [[CrossRef](#)]
33. Huang, Y.L. Computer-Aided Diagnosis Using Neural Networks and Support Vector Machines for Breast Ultrasonography. *J. Med. Ultrasound* **2009**, *17*, 17–24. [[CrossRef](#)]
34. Nascimento, C.D.L.; Silva, S.D.D.S.; Silva, T.A.D.; Pereira, W.C.D.A.; Costa, M.G.F.; Costa Filho, C.F.F. Breast Tumor Classification in Ultrasound Images Using Support Vector Machines and Neural Networks. *Res. Biomed. Eng.* **2016**, *32*, 283–292. [[CrossRef](#)]
35. Kumar, A.; Saini, R.; Kumar, R. A Comparative Analysis of Machine Learning Algorithms for Breast Cancer Detection and Identification of Key Predictive Features. *Trait. Du Signal* **2024**, *41*, 127. [[CrossRef](#)]
36. Chatterjee, D.; Ghosh, P. Comparative Analysis of Machine Learning Algorithms for Breast Cancer Classification: SVM Outperforms XGBoost, CNN, RNN, and Others. *bioRxiv* **2024**, *41*, 127–140. [[CrossRef](#)]
37. Mendelson, E.B.; Böhm-Velez, M.; Berg, W.A.; Whitman, G.J.; Feldman, M.I.; Madjar, H.; Rizzatto, G.; Baker, J.A.; Zuley, M.; Stavros, A.T.; et al. ACR BI-RADS Ultrasound. In *ACR BI-RADS Atlas, Breast Imaging Reporting and Data System*; American College of Radiology: Reston, VA, USA, 2013.
38. Belongie, S.; Malik, J.; Puzicha, J. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 509–521. [[CrossRef](#)]
39. Stone, M. Cross-validatory choice and assessment of statistical predictors. *J. R. Stat. Soc. B* **1974**, *36*, 111–147. [[CrossRef](#)]
40. Yang, H.; Yang, D. CSwin-PNet: A CNN-Swin Transformer Combined Pyramid Network for Breast Lesion Segmentation in Ultrasound Images. *Expert Syst. Appl.* **2023**, *213*, 119024. [[CrossRef](#)]
41. Sun, S.; Cao, Z.; Liao, D.; Lv, R. A Magnified Adaptive Feature Pyramid Network for Automatic Microaneurysms Detection. *Comput. Biol. Med.* **2021**, *139*, 105000. [[CrossRef](#)]
42. Alam, M.S.; Wang, D.; Sowmya, A. AMFP-net: Adaptive Multi-Scale Feature Pyramid Network for Diagnosis of Pneumoconiosis from Chest X-Ray Images. *Artif. Intell. Med.* **2024**, *154*, 102917. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.