

Comparison of Ultrasound Image Classification Methods for Benign and Malignant Breast Tumors Based on Texture and Shape Characteristics

Nova Senandung, Nanda Firdayana*, Dewi Anggun Puspita Septiani, Syahwa Ais Saputri, Heni Sumarti

Departement of Physics, Faculty of Science and Technology, Universitas Islam Negeri Walisongo Semarang, Indonesia

Article Info

Article History:

Received:
18 December 2025

Accepted:
18 February 2026

Published:
15 May 2026

Keywords:

Breast ultrasound, Texture and shape features, Random Forest, Naïve Bayes, K-fold cross-validation

Abstract

Early detection of breast tumors is important to accurately distinguish benign and malignant lesions. This study aims to compare the performance of Random Forest and Naïve Bayes algorithms in the classification of breast ultrasound images based on texture and shape features. The dataset comes from BUSI which consists of two classes: benign 891 images and malignant 421 images, with images through the pre-processing stage, histogram feature extraction, GLCM, as well as morphological features (circularity and elongation). Feature rankings using Relief show that GLCM Homogeneity has the greatest contribution in distinguishing the two classes. Performance evaluation was carried out using K-Fold Cross Validation with variations of K=5, 10, 15, 20, and 25. The results showed that consistently placed Random Forest as the best-performing model. Random Forest achieved the highest accuracy at k-fold-5 at 74.18%, with stable AUC values at 0.771-0.777, sensitivity reaching 78.25%, and better specificity (62-63%) across the fold. In contrast, Naïve Bayes showed lower accuracy with a maximum value of 59% at k-fold-25, AUC in the range of 0.68, and low specificity (42-43%) despite the relatively high sensitivity. These findings confirm that across all k-fold validations, Random Forest remains the most balanced and reliable model for distinguishing benign and malignant in breast.

INTRODUCTION

Breast cancer is one of the biggest health problems for women in the world and is the second highest cause of cancer death, after lung cancer Sung et al. (2021). Based on Globocan 2020 data, there are 68,858 new cases of breast cancer in Indonesia, or around 16.6% of all cancer cases. This number shows a significant increase from year to year, so accurate early detection is an important aspect in efforts to reduce mortality.

In general, breast tumors are divided into two types, namely benign and malignant. Benign tumors are usually small, round, soft lumps, and are not cancerous because their growth does not spread or damage the surrounding tissue. On the other hand, malignant tumors are characterized by an asymmetrical shape, rough surface, and often painful sensations (Hidayat & Davi, 2025). Early detection is an important step to reduce the mortality rate, in addition to avoiding various risk factors. One commonly used method is breast ultrasound (mammary ultrasound), which is a special examination to look at the condition of breast tissue and detect abnormalities such as cysts or tumors (Hidayat & Davi, 2025).

Ultrasound imaging is one of the medical imaging methods that is often used to detect and diagnose breast tumors because it is non-invasive, safe, relatively affordable, does not use ionizing radiation and can be widely used in a variety of clinical conditions. For example, a study by Donners et al. (2021) suggests that ultrasound can be an effective adjunctive modality in distinguishing malignant and benign lesions, although the results are still highly dependent on the skills of the operator and radiological interpretation. However, ultrasound imaging faces various technical constraints, such as high speckle noise levels, low contrast between normal tissue and tumor tissue, and less clear lesion boundaries. This condition makes the automatic segmentation process difficult, even though this stage is very important in the analysis of medical images to accurately determine the location, size, and shape of the tumor.

To overcome these limitations, various efforts have been made through the development of computer-based diagnostic assistance systems or *Computer-Aided Diagnosis (CAD)*. This system utilizes digital image processing and machine learning to recognize certain patterns in ultrasound images. One of the widely used approaches is the extraction of features that present the visual characteristics of the tumor. The two most commonly used types that have proven to be effective are texture and shape features. Texture traits are able to capture patterns of intensity and granular structure in images, for example *Gray-Level Co-occurrence Matrix (GLCM)*, *local binary pattern (LBP)*, or *histogram of oriented gradients (HOG)* which proves useful for distinguishing benign and malignant lesions (Wei et al., 2020). Meanwhile, shape/morphology features highlight contours, smoothness of borders, aspect ratios, eccentricity, and variability of lesion shape because clinically malignant lesions generally have uneven borders, irregular shapes and different orientations compared to benign lesions Byra et al. (2019). Thus the combination of texture and shape becomes a logical strategic to improve classification performance in the context of US breast tumors.

As technology evolves *machine learning*, various algorithms have been applied for the classification of medical images including ultrasound images of the breast. Classical algorithm Naïve Bayes and Random Forest have been widely applied to ultrasound images of breast tumors, with the advantage of interpretability and lower data requirements compared to deep learning. Random Forest is known to be resilient to overfitting and capable of handling non-linear features, while Naïve Bayes offers a simple and interpretable probabilistic baseline. Like a Study by Wei et al. (2020) shows that classical algorithms such as Random Forest and Naïve Bayes provide competitive accuracy compared to deep learning models in benign vs malignant ultrasound datasets, with Random Forest showing the best performance among classical methods. Nonetheless, few specific studies have compared these two methods with an explicit focus on texture and shape features on ultrasound images of breast tumors in consolidated studies. The gap in this study shows that it is necessary to carry out a systematic

analysis that compares these classical methods (Random Forest and Naive Bayes) in the context of the extraction of texture and shape features in ultrasound images.

Therefore, this study aims to compare the classification performance between Naïve Bayes and Random Forest in exposing benign and malignant breast tumor lesions based on texture and shape characteristics extracted from ultrasound images. By conducting a comprehensive evaluation of the texture and morphological features and implementation of the two algorithms, it is hoped that this study can contribute to the CAD literature on breast ultrasound and increase the potential for clinical applications in the early detection and diagnosis of breast tumors.

METHOD

This study uses breast ultrasound imaging data compiled by Khaled and Fahmy (2023), taken from Breast Ultrasound Images (BUSI) available on Kaggle. The dataset contains three classes of ultrasound images, namely normal, benign, and malignant. However, this study only focused on two categories namely benign and malignant, since the main objective was to distinguish between benign and malignant tumors on ultrasound images. The total data used amounted to 647 images, consisting of 437 benign images and 210 malignant images. Nonetheless, some images were excluded from the analysis process because their quality did not meet the eligibility criteria. Images with codes 207, 272, 338, 429, 431, and 270 for the class *benign* and 160 images for the class *malignant* deleted as a result *Noise* high, low resolution, or inconspicuous lesion areas that may affect feature extraction results (Liberini et al., 2020).

After the data selection and filtering process is carried out, all remaining images are then prepared through the pre-processing stage to ensure quality and consistency before entering the feature extraction process. This process aims to reduce *noise*, standardize the size, and highlight important areas of the image for more accurate analysis. All stages are done using images in PNG format obtained from Kaggle.

The pre-processing steps are carried out systematically as follows. First, all images are converted to 8-bit grayscale format to standardize color intensity and reduce computational complexity. Second, noise reduction is carried out using *Median Filter* 3×3 in size to reduce *speckle noise* that often appears on ultrasound images Gonzalez & Woods (2018). Third, the entire pixel intensity value is normalized into the range [0, 1] to minimize the scale difference between images.

Furthermore, ROI cropping is carried out automatically using *mask* built-in dataset to ensure only areas containing tumor lesions are analyzed. Finally, the entire image was converted to a size of 224×224 pixels based on a study by (Simonyan & Zisserman (2015), so that the entire dataset has uniform dimensions (Behmann et al., 2018).

With the image having gone through the pre-processing stage and having uniform quality and size, the next process is feature extraction or feature extraction that represents the visual characteristics of the tumor. In this study, the extracted features are divided into two major groups, namely texture features and shape features (Sharma & Abrol, 2020).

Texture features are used to depict the distribution patterns and spatial relationships between pixels in the ultisographic image Materka & Strzelecki (1998). In this study, texture features were extracted using first-order statistics from the intensity histogram as well as the Gray Level Co-occurrence Matrix (GLCM). First-order histograms provide information regarding the distribution of pixel intensity without considering spatial proximity. The value of the intensity probability can be expressed as $p(i) = \frac{n_i}{N}$, with n_i is the number of pixels at the intensity level to - and N is the total number of pixels in the image. From this histogram, some basic statistics are calculated, including: i (Krisdiyanto & Sumarti, 2024)

- a. Mean

$$m = \sum_{i=1}^{L-1} i \cdot p(i) \quad (1)$$

b. Standard Deviation

$$\sigma = \sqrt{\sum_{i=0}^{L-1} (i - m)^2 p(i)} \quad (2)$$

c. variasi

$$\sigma^2 = \sum_{i=0}^{L-1} (i - m)^2 p(i) \quad (3)$$

d. Entropy

$$Entropy = - \sum_{i=0}^{L-1} \log^2(p(i)) \quad (4)$$

e. Skewness

$$Skewness = \sum_{i=0}^{L-1} (i - m)^3 p(i) \quad (5)$$

f. Kurtosis

$$Kurtosis = \sum_{i=0}^{L-1} (i - m)^4 p(i) \quad (6)$$

These statistics represent the character of the intensity distribution so that they are able to distinguish the level of brightness, smoothness, and randomness of the network structure.

In addition to the histogram, texture features are also calculated using GLCM that captures the spatial relationships between pixels of a certain intensity at a certain distance and orientation. If it is the probability of the occurrence of an intensity pair and in the GLCM, then the features calculated include $P(i, j) i j$ (Krisdiyanto & Sumarti, 2024).

a. Contrast

$$Contrast = \sum_i \sum_j (i - j)^2 p(i, j) \quad (7)$$

b. Correlation

$$Correlation = \frac{1}{\sigma_x \sigma_y} \sum_i = 1 \sum_j = 1 (1 - \mu_x)(1j - \mu_j)p(i, j) \quad (8)$$

c. Energy

$$Energy = \sum_{i,j=0}^{N-1} p(i, j)^2 \quad (9)$$

d. Homogeneity

$$Homogeneity = \sum_i \sum_j \frac{p(i,j)}{1 + (i - j)} \quad (10)$$

The resulting features of GLCM can describe the degree of contrast, uniformity, and regularity of texture patterns in breast tissue. In general, malignant tissues show a higher level of heterogeneity than benign tissues, so the difference in the statistical value of these textures can be used as an important indicator in the classification process. The selection of the GLCM method is based on its ability to capture local variations of pixel intensity that reflect network heterogeneity. For example, a high contrast value indicates a sharp difference in intensity between pixels, while *homogeneity* represents the uniformity of the network structure of Zhang et al. (2020). This characteristic is relevant because malignant tissues tend to have a more non-uniform structure than benign tissues.

In addition to texture features, this study also extracted shape features (*shape features*) to represent the geometric characteristics of the tumor mass. The two main parameters used are *circularity* and *elongation*. Value *circularity* calculated using an equation that is referenced from (Krisdiyanto & Sumarti, 2024)

$$C = \frac{4\pi A}{P^2} \quad (11)$$

with A as the area of the object and P as the circumference of the object, which represents the degree of roundness of the tumor mass. Meanwhile, Elongation is calculated as the ratio between the length of the major and minor axes of an object. Benign masses generally have high circularity values and low elongation because they are more regular in shape and close to round. In contrast, malignant masses tend to have a more oval, asymmetrical, and irregular shape. The combination of texture features and shape features provides a more comprehensive visual description of the mass character of the tumor, both in terms of its internal and external patterns.

Once all features have been extracted, the next stage is feature ranking to determine the contribution of each feature in distinguishing benign and malignant tumors. The ranking process was carried out using WEKA 3.8 with the Relief Attribute Evaluator – Ranker method. It is important to affirm that Relief is not Information Gain and does not use the entropy formula. Relief works by evaluating how well a feature is able to distinguish instances between classes based on the nearest hit and nearest miss. (As studied by (Li et al., 2016)(Li et al., 2016)Li et al., 2016) Those that provide a big difference between classes will get high weight, while features that don't affect will get low weight. Thus, the Relief method focuses more on the ability of features to separate classes based on the proximity of values, rather than based on entropy theory as in Information Gain or Gain Ratio. The Relief method provides weight $W(A)$ for an instance and his classmates (*Nearest Hit*) as well as other class neighbors (*Nearest Miss*).

In general, feature weights are updated after processing each $W(A)$ instance R (Abdulkareem, 2020) as follows:

$$W(A) = W(A) - \frac{\text{diff}(A, R, H)}{m} + \frac{\text{diff}(A, R, M)}{m}$$

In this formulation, $W(A)$ represents the weight of feature A , while m denotes the number of randomly selected instances used in the calculation process. The variable R refers to the instance currently being evaluated, H is the Nearest Hit, defined as the closest instance belonging to the same class as R , and M is the Nearest Miss, which corresponds to the closest instance from a different class. The function $\text{diff}(A, I1, I2)$ is employed to quantify the difference in the value of feature A between two instances $I1$ and $I2$. Consequently, the adjustment of feature weights reflects the capability of each feature to distinguish between instances from the same class and those from different classes.

Features that provide a large difference between classes (i.e. large and small) will gain high weight, while features that have no effect will receive low weight. Thus, the Relief method focuses more on the ability of features to separate classes based on the proximity of values, rather than on the theory of entropy. $diff(A, R, M)diff(A, R, M)$

With Relief ratings, the features that have the greatest contribution to class differentiators can be identified more accurately before they are used at the next stage of classification.

Although this study used the Relief Attribute Evaluator, it is important to note that other ranking methods, such as those based on information theory (e.g. Information Gain or Gain Ratio), are also available. However, Relief is better suited because it directly measures feature quality based on the separation of spacing between classes, making it a top choice in this study.

Once important features have been identified through the grading process, the next step is to implement these features into the classification model to assess the ability of each algorithm to recognize the visual pattern of the tumor. Image classification was carried out using two comparison algorithms, namely Random Forest (RF) and Naïve Bayes (NB) using WEKA 3.8 software. The selection of these two methods is based on significant differences in approach: RF is *ensemble* and decision tree-based, whereas NB is probabilistic-based.

Random Forest works by building a large number of decision trees using the *bootstrap sampling*. Each tree is trained with a different subset of data and a subset of features to reduce risk *overfitting*. The final classification results are determined through the mechanism *majority voting* of all the trees that were built. This approach has been proven to produce a stable, resistant model to *Noise*, and has high accuracy on multi-dimensional datasets (Breiman, 2001).

The Random Forest method is used for the image classification process Sheykhmousa et al. (2020). Random Forest is one of the algorithms *machine learning* who belong to the *ensemble learning*, where multiple decision trees are built and then combined to produce a final decision based on the mechanism *Voting* mayoritas.

In addition to being used as a classifier, Random Forest is also able to rank features based on their level of contribution to the classification process. The importance of a feature is evaluated using Gini Impurity, which is a measure that describes the level of impurity of a node in the decision tree (Bickel et al., n.d.). The smaller the Gini value, the better the feature is at separating classes. The Gini Impurity value is calculated using an equation (Smys et al., 2019):

$$GiniImpurity = 1 - \sum_{j=1}^n (p_j)^2 \quad (17)$$

The Gini Impurity value is determined using Equation (13), where n designates the quantity of target classes, j indicates the class index, and p_j reflects the proportion of samples within the j -th class. This measure serves to assess the homogeneity of data contained within a specific node; a lower Gini Impurity value signifies superior class separation.

The classification process using the Random Forest algorithm was carried out through two main stages, namely the model training stage with a *training set* and the evaluation stage using the *cross-validation* method. In the *cross-validation* process, the dataset is divided into several subsets or *folds* that are alternately used as test data to assess the performance of the model. Performance evaluation is carried out using the *K-Fold cross-validation* method, where the data is divided into K groups (*folds*) and tests are carried out on each group so that a more stable and representative performance value is obtained.

On the other hand, the Naïve Bayes method (NB) is based on Bayes' Theorem assuming that each feature is mutually independent (Rish, I., 2001).

The probability of a class is then calculated using the following equation:

:

$$P(Y | X) = \frac{P(X | Y)P(Y)}{P(X)} \quad (18)$$

with (Y) as the class (benign or malignant) and (X) as the vector of the image feature. To avoid zero probability, the Laplace Smoothing technique is applied. This method shows advantages in terms of computing speed as well as good performance on medium-sized datasets with relatively independent features. Once the classification process is carried out, a validation stage is required to ensure that the performance of the resulting model is completely accurate and does not depend on specific data sharing.

To ensure the reliability of the developed model, this study applied the *K-Fold Cross-Validation* technique with variations in K-values of 5, 10, 15, 20, and 25 Kohavi (1995). In this method, the dataset is randomly divided into K subsets of relatively equal size. Each subset was alternately used as test data, while the other subset was used as training data. The procedure is repeated until each data is used as test data exactly once. This cross-validation *approach* was chosen because it was able to provide a more stable model performance estimate compared to the single-division (*hold-out*) method. In addition, the use of K-value variations allows observation of the consistency of model performance on the different proportions of training data and test data, while helping to reduce the risk of *overfitting*. The validation results are then used as a basis for conducting a more in-depth analysis of model performance.

The evaluation of the model's performance is carried out by calculating four main parameters (Krisdiyanto & Sumarti, 2024), namely:

1. Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (19)$$

2. Recall

$$Recall = TPR = \frac{TP}{TP + FN} \quad (20)$$

3. Specificity (true negative rate/ TNR)

$$Spesifitas = TNR = \frac{TN}{TN + FP} \quad (21)$$

4. Accuracy

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (22)$$

5. AUC (Area Under Curve)

AUC is the area under the ROC curve (TPR vs FPR). The AUC value indicates the model's ability to distinguish between positive and negative classes

$$AUC = \int_0^1 TPR(FPR) d(FPR) \quad (23)$$

(Pawłowska et al., 2023)

By general interpretation is shown in Table 1

Table 1 Interpretation of AUC Values on the Quality of Classified Models

	Model Quality (General Interpretation)	Remarks
AUC = 0.5	Random Guess	Model performance is no better than guessing the outcome at random.
AUC < 0.7	Bad	The model has a low discriminating ability.
AUC 0.7 – 0.8	Acceptable	The model has adequate discriminating capabilities.
AUC 0.8 – 0.9	Excellent	Models have strong discriminating abilities.
AUC > 0.9	Excellent (Outstanding)	The model has an incredible and almost perfect discriminating ability.
AUC = 1.0	Perfect	The model can distinguish all positive and negative cases perfectly.

(Guetari et al., 2023)

RESULTS AND DISCUSSION

The data used in this study were in the form of breast ultrasound images which were grouped into two categories, namely benign and malignant. Each image has different resolutions and varying noise levels, which are affected by the type of device and the conditions when shooting. Therefore, the digital image processing stage is needed to standardize the visual quality of the image so that the process of extraction and feature analysis can be carried out more optimally.

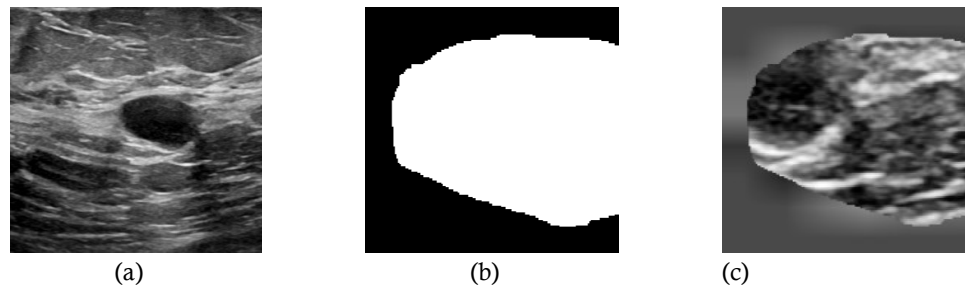


Figure 1. Ultrasound Image of Benign Lesions on (a) Original, (b) ROI Masks, and (c) ROI For Features

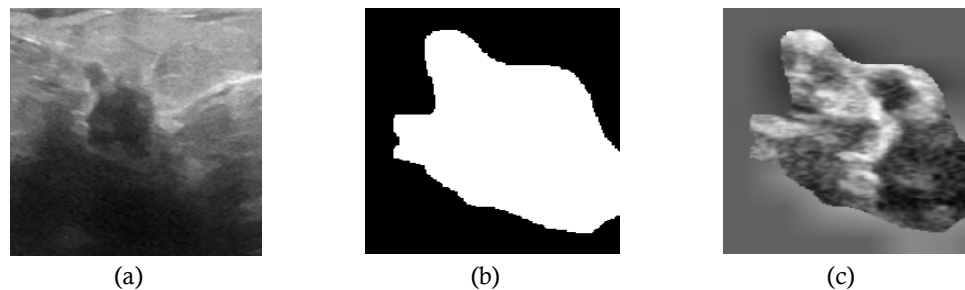


Figure 2. Ultrasound Images of Malignant Lesions on (a) Original, (b) ROI Masks, and (c) ROI For Features

On the image *original* benign lesions shown in Figure 1(a), the grayish pattern appears relatively homogeneous with the shape of the lesions tending to be round and regular, as well as fairly

clear boundaries to the surrounding tissues. On the contrary, the image *original* the malignant lesions in Figure 2(a) show a coarser texture, uneven distribution of intensity, and irregular lesion contours so that the boundary between the lesion and the surrounding tissue appears less pronounced. At the stage *ROI masks*, the benign lesions in Figure 1(b) show a more symmetrical shape with smooth margins, while the malignant lesions in Figure 2(b) have an asymmetrical shape, many curves, and margins that tend to be irregular. Next, at the stage *ROI for features*, the benign image in Figure 1(c) shows a more uniform texture with relatively low intensity variation and consistent internal structure. In contrast, the malignant imagery in Figure 2(c) shows a denser, random, and heterogeneous texture with higher contrast levels. Overall, comparisons at each stage showed that benign lesions generally had more regular and homogeneous characteristics, whereas malignant lesions displayed more complex, irregular patterns, and had greater textural variation (Wei et al., 2020).

The feature extraction process is carried out in the (ROI) by utilizing two main feature groups, namely texture features obtained from intensity histograms and (GLCM), as well as shape features resulting from the segmentation stage. The average values of each of the texture and shape features for both types of tumors are presented in Table 2.

Table 2. Average Statistics of Texture and Shape Features

Texture Features	Malignant	Benign	Ranked Attributes
Mean	104. 1544 $\pm$ 24. 338	115. 3872 $\pm$ 31. 4611	0.00892
Standard Deviation	42. 8082 $\pm$ 8. 8566	41. 0072 $\pm$ 9.8606	0.01254
Variance	1910.8847 $\pm$ 824. 2608	1778. 9331 $\pm$ 856. 9176	0.01051
Entropy	10. 7226 $\pm$ 0. 0666	10. 7399 $\pm$ 0.0663	0.00237
Skewness	0.9075 $\pm$ 0.7047	0.5372 $\pm$ 1.2393	0.00237
Kurtosis	2. 4859 $\pm$ 4. 6969	3.3495 $\pm$ 10.3948	0.00129
GLCM_Contrast	211. 8942 $\pm$ 91. 9453	181. 0747 $\pm$ 125. 3497	0.01532
GLCM_Correlation	0.9434 $\pm$ 0. 0161	0. 9489 $\pm$ 0.0267	0.00958
GLCM_Energy	0.1915 $\pm$ 0.082	0.2387 $\pm$ 0.1216	0.01393
GLCM_Homogeneity	0.4842 $\pm$ 0.0918	0.5378 $\pm$ 0. 1355	0.0217
Circularity	0.1733 $\pm$ 0.1932	0. 2908 $\pm$ 0. 2817	0.00476
Elongation	0.5224 $\pm$ 0. 239	0. 4096 $\pm$ 0. 2853	0.00609

The results presented in Table 2 show a significant difference between benign and malignant images. The mean pixel intensity in malignant tumors (104.15) was lower compared to benign (115.38), reflecting the hypochoic characteristics commonly found in malignant lesions. Higher *standard deviation* and *variance* values in malignant indicate greater variation in intensity, in line with the high degree of malignant tissue heterogeneity due to aggressive growth patterns. In addition, *the entropy* and *kurtosis* parameters indicate that the intensity distribution in malignant images is more random and extreme, while *higher skewness* values indicate the presence of pixel distribution asymmetry. In the GLCM feature analysis, the *contrast* value in the malignant image (211.89) was greater than in the benign (181.07), indicating a higher level of texture irregularity. In contrast, *the energy* and *homogeneity* values are higher in benign images, reflecting a relatively smoother and uniform texture. The *correlation* value is also slightly greater in benign, indicating a more consistent inter-pixel relationship compared to malignant tissue.

The shape feature also shows a fairly consistent difference between the two classes. Value *circularity* The benign image (0.29) was higher than the malignant (0.17), reflecting the shape of the lesion which tends to be more rounded and regular. On the other hand, the value *elongation* larger in malignant images (0.52), indicating that the lesions have a more elongated, asymmetrical, and

irregular shape. These findings are in line with the morphological characteristics of malignant tumors which are generally invasive. Based on the results of the analysis *ranking attributes*, GLCM features *homogeneity* ranks highest with a value of 0.0217, which indicates that this feature has the strongest discriminating ability in distinguishing benign and malignant images. Value *homogeneity* The low one is related to the increasing irregularities of the tissue structure, so it is often used as an indicator of malignancy. These results are consistent with the findings Prabusankarlal et al. (2017) who report that *inverse difference moment* (IDM) or homogeneity is one of the most effective features in the classification of breast mass. Thus, the low homogeneity value in the malignant image reflects a rough, random, and unstructured texture pattern.

The classification process in this study was carried out by applying the Random Forest and Naïve Bayes algorithms using a combination of texture features and shape features as model input. The performance of the two algorithms was evaluated through *the k-fold cross-validation method* with variations in the number of *folds* as many as 5, 10, 15, 20, and 25. A summary of the evaluation results for each test scheme is presented in Table 3.

Table 3. Results of Classification Performance Evaluation Using Random Forest and Naïve Bayes

	FOLD	TN	FN	FP	HCMC	AUC	AC	SENSITIVITY	SPECIFICITY
Random Forest	5	107	102	63	367	0.7732	74.18	78.25	62.94
	10	102	107	66	364	0.7745	72.93	77.28	60.71
	15	100	109	61	369	0.7756	73.4	77.2	62.11
	20	98	111	58	372	0.7768	73.55	77.02	62.82
	25	101	108	64	366	0.7714	73.08	77.22	61.21
Naïve Bayes	5	172	37	235	195	0.6822	57.43	84.05	42.26
	10	171	38	235	195	0.6824	57.28	83.69	42.12
	15	174	35	231	199	0.6853	58.37	85.04	42.96
	20	173	36	230	200	0.6883	58.37	84.75	42.93
	25	176	33	229	201	0.6874	59	85.9	43.46

Overall, Random Forest showed the most consistent performance across the entire fold count scheme used. The AUC value is in the range of 0.771 to 0.776 which indicates the model's ability to distinguish between the two classes. The sensitivity value reached the highest level of 78.25%, indicating that most malignant cases could be correctly identified by the model. Meanwhile, the specificity value was in the range of 62–63%, which indicates that although the model is quite effective in detecting malignant cases, there is still a chance of misclassification in benign cases. Random Forest's superior performance is inseparable from the *ensemble learning* principle that combines decisions from multiple decision trees, so that the model is able to represent the relationships between complex and nonlinear features, as are commonly found in medical imaging.

In contrast, the Naïve Bayes method results in relatively lower performance, with an accuracy rate in the range of 57–59% and an AUC value of around 0.68. Although the sensitivity value obtained is quite high, which is around 84–85%, the specificity value is relatively low, which is only 42–43%. This suggests that models tend to classify benign images as malignant. This performance is related to the basic assumption of Naïve Bayes which considers each feature to be mutually independent, while in ultrasound imaging the relationship between texture features and shapes is generally complex and interdependent.

The difference in performance generated by the two methods suggests that Random Forest is more suitable for the classification of medical images that have a nonlinear distribution of intensity as well as complex interdependencies. Although Naïve Bayes has advantages in terms of speed and simplicity, this method is less able to represent the complexity of the data characteristics. Overall,

Random Forest can be judged as a more reliable approach in distinguishing benign and malignant images based on a combination of texture and shape features. These findings support the development of a more accurate *Computer-Aided Diagnosis* (CAD) system for the detection of breast cancer based on ultrasound images.

A number of previous studies have developed a further approach in the classification of breast ultrasound images. Study by Wei et al. (2020) for example, combining texture features such as GLCM, LBP, and HOG with morphological features through the *score-level fusion* of the two classifiers, which have been proven to be able to improve the performance of the classification. Meanwhile, research by Daoud et al. (2020) Blending features *deep learning* VGG19 based with *handcrafted* such as GLCM and morphological features, thus obtaining higher accuracy than conventional approaches. Other research also by Ieniewski (2020) extracted GLCM features directly from the raw RF signal and applied them to a logistic regression model, with a reported AUC value of 0.83. Approach *meta-learning* has also been introduced by Byra et al. (2022) Through the development of a Meta network architecture that adaptively selects the most appropriate classification model between texture features and shape features based on image characteristics. In addition, Benli (2024) Propose a Method-Based Method *two-dimensional empirical mode decomposition* (biEMD) to extract multi-subband features, such as GLCM, LBP, and HOG, and report significant improvements in classification performance.

This research has a number of advantages that support its contribution to the development of an ultrasound image-based early detection system for breast cancer. The method applied is able to distinguish benign and malignant ultrasound images with a good level of performance, so that it has the potential to help medical personnel in the initial diagnosis process. In addition, this study combines texture features and shape features simultaneously, so that the characteristics of the lesions can be represented more thoroughly than studies that use only one type of feature. The utilization of the Random Forest algorithm has also been proven to provide stable and effective classification performance on medical data that has complex feature relationships. Overall, the results of this study can be a methodological basis for the development of an *ultrasound image-based Computer-Aided Diagnosis* (CAD) system in future research.

Although this study shows quite good performance results, there are still some limitations that need to be considered. The ultrasound image datasets used were relatively limited and came from only one source, so the variation in image characteristics has not been fully accommodated. In addition, the classification system developed is not yet equipped with a user interface (GUI) that supports direct application in a clinical environment. The study also only evaluated two classification algorithms based on conventional methods, namely Random Forest and Naïve Bayes, without considering other approaches such as *deep learning* or more complex ensemble methods. Another limitation is that generalization testing has not been carried out using datasets from various institutions, so the performance of the model in different clinical conditions cannot be ascertained.

Further research can be directed to some further development. One of them is the application of the *stacking ensemble method* which has the potential to increase the level of accuracy while maintaining the stability of model performance. In addition, the utilization of *deep learning approaches* such as *Convolutional Neural Networks* (CNN) or hybrid models can be considered to produce a more automated and representative feature extraction process. The scope of research can also be expanded by involving datasets from various sources so that the variation of images becomes more diverse, so that the generalization ability of the model can be increased. On the implementation side, the development of an interactive user interface (GUI) is required for the CAD system to be applied directly to clinical devices. Further evaluation in a real clinical environment is also important to determine the performance of the system under actual conditions of use.

CONCLUSION

This study shows that texture and shape features are able to effectively distinguish images of benign and malignant breast tumors. The GLCM Homogeneity feature proved to be the most discriminating parameter, with lower values in malignant lesions reflecting rougher and irregular textures. The classification results show that Random Forest performs best with an AUC of 0.7714–0.7768, an accuracy of about 73%, a sensitivity of up to 78.25%, and a specificity of 62–63%. In contrast, Naïve Bayes has lower accuracy and specificity. Overall, the combination of texture features, shape features, and Random Forest's algorithm has strong potential to be implemented in Computer-Aided Diagnosis (CAD) systems to improve the detection of breast lesions in ultrasound images.

REFERENCE

- Abdulkareem, M. I. Y. M. H. (2020). Pilgrims tracking and monitoring based on IoT You. *Journal of Physics: Conference Series PAPER*. <https://doi.org/10.1088/1742-6596/1530/1/012015>
- Behmann, J., Bohnenkamp, D., Paulus, S., & Mahlein, A. (2018). Spatial Referencing of Hyperspectral Images for Tracing of Plant Disease Symptoms. *Journal of Imaging*, 4(143). <https://doi.org/10.3390/jimaging4120143>
- Benli, Ş. G. (2024). A Novel Approach for Breast Ultrasound Classification Using Two-Dimensional Empirical Mode Decomposition And Multiple Features. *Experimental Biomedical Research*, 7(1), 27–38.
- Bickel, P., Diggle, P., Fienberg, S., Gather, U., Olkin, I., & Zeger, S. (n.d.). *Springer Series in Statistics*.
- Breiman, L. E. O. (2001). Random Forests. 5–32.
- Byra, M., Galperin, M., & Comstock, C. (2019). Breast Mass Classification in Sonography With Transfer Learning Using A Deep Convolutional Neural Network And Color Conversion. *Medical Physics*, 46(2), 746–755. <https://doi.org/10.1002/mp.13361>
- Byra, M., Karwat, P., Ryzhankow, I., Komorowski, P., Klimonda, Z., Fura, L., Pawlowska, A., Zolek, N., & Litniewski, J. (2022). Deep meta-learning for the selection of accurate ultrasound based breast mass classifier.
- Daoud, M. ., Samir Abdel-Rahman, T. M. B., Mahasen S. Al-Najar, F. H. A.-H., & Alazrai, R. (2020). Breast Tumor Classification in Ultrasound Images Using Combined Deep and Handcrafted Features. *Sensors*, 1–20. <https://doi.org/10.3390/s20236838>
- Donners, R., Gutzeit, A., Gehweiler, J. E., Manneck, S., Kovacs, K., & Harder, D. (2021). Orthopaedic surgeons do not consult radiology reports . Fact or fiction ? *European Journal of Radiology*, 142(July), 109870. <https://doi.org/10.1016/j.ejrad.2021.109870>
- Gonzalez, R. C., & Woods, R. E. (2018). *Digital Image Processing*, 4e.
- Gueteri, R., Ayari, H., & Sakly, H. (2023). Computer-aided diagnosis systems : a comparative study of classical machine learning versus deep learning-based approaches. *Knowledge and Information Systems*. <https://doi.org/10.1007/s10115-023-01894-7>
- Hidayat, R., & Davi, M. (2025). Hierarchical Ensemble Learning untuk Klasifikasi Tumor Payudara pada Citra Ultrasonografi (USG). 8(1), 48–53.
- Ieniewski, M. A. N. (2020). Study of Classification of Breast Lesions Using Texture GLCM Features Obtained From The Raw Ultrasound Signal. *Image Anal Stereol*, 39, 129–145. <https://doi.org/10.5566/ias.2113>

- Khaled, H., & Fahmy, A. (2023). Letter to the Editor . Re : “[Dataset of breast ultrasound images by W . Al-Dhabyani , M . 48. <https://doi.org/10.1016/j.dib.2023.109247>
- Krisdiyanto, R., & Sumarti, H. (2024). Classification of mammographic image based on texture features with Random Forest method for identification of breast tumors. *Journal of Holistic Medical Technologies*, 1(1), 1–9.
- Li, M., Gao, Y., & Sang, N. (2016). Exploiting Learnable Joint Groups for Hand Pose Estimation. *The Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21)* Exploiting.
- Liberini, V., Huellner, M. W., Grimaldi, S., Finessi, M., Thuillier, P., Muni, A., Pellerito, R. E., Papotti, M. G., Piovesan, A., Arvat, E., & Deandreis, D. (2020). The Challenge of Evaluating Response to Peptide Receptor Radionuclide Therapy in Gastroenteropancreatic Neuroendocrine Tumors: The Present and the Future. *Diagnostics*, 10(1083), 1–31.
- Materka, A., & Strzelecki, M. (1998). *Texture Analysis Methods – A Review*.
- Pawłowska, A., Karwat, P., & Żołek, N. (2023). Letter to the Editor. Re: “[Dataset of breast ultrasound images by W. Al-Dhabyani, M. Gomaa, H. Khaled & A. Fahmy, Data in Brief, 2020, 28, 104863].” *Data in Brief*, 48. <https://doi.org/10.1016/j.dib.2023.109247>
- Prabusankarlal, K. M., Thirumoorthy, P., & Manavalan, R. (2017). Classification Of Breast Masses In Ultrasound Images Using Self-Adaptive Differential Evolution Extreme Learning Machine And Rough Set Feature Selection. *Journal of Medical Imaging*, 4(2). <https://doi.org/10.1117/1.JMI.4.2.024507>
- Sharma, R., & Abrol, P. (2020). Image Feature Extraction Techniques. *International Journal of Scientific and Technical Advancements*, 6(3), 125–128.
- Sheykhoumoussa, M., Mahdianpari, M., & Ghanbari, H. (2020). Support Vector Machine vs . Random Forest for Remote Sensing Image Classification : A Meta-analysis and systematic review. *JSTARS*. <https://doi.org/10.1109/JSTARS.2020.3026724>
- Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *Conference Paper at ICLR 2015*, 1–14.
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020 : GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. 71(3), 209–249. <https://doi.org/10.3322/caac.21660>
- Wei, M., Du, Y., Wu, X., Su, Q., Zhu, J., Zheng, L., Lv, G., & Zhuang, J. (2020). A Benign and Malignant Breast Tumor Classification Method via Efficiently Combining Texture and Morphological Features on Ultrasound Images. *Computational and Mathematical Methods in Medicine*, 2020. <https://doi.org/10.1155/2020/5894010>