European Journal of Molecular & Clinical Medicine (EJMCM)
ISSN: 2515-8260 Volume 10, Issue 04, 2023

BREAST CANCER CLASSIFICATION

Abdul Khayyum Farooqui
Research Scholar
Sreenidhi Institute of Science and Technology
abdulkhayyum.519@gmail.com

Dr. Preethi Jeevan
Associate Professor
Sreenidhi Institute of Science and Technology
preethij@sreenidhi.edu.in

Abstract

Breast cancer detection is the initial stage of a cancer diagnosis. Therefore, more accurate classifiers are always preferred. A high accuracy classifier offers an extremely low probability of incorrectly classifying a cancer patient. This research investigates the performance of a modified and improved version of the guiding hypothesis of the logistic regression. Gradient descent and complex optimization techniques are both used to minimise the cost function. The weighting factor for the hypothesis, which is a sigmoid function, is determined by the number of features, the size of the dataset, and the kind of optimization technique used, as can be seen. By carefully choosing the value, which relies on the number of features and the kind of optimization methods utilised, the accuracy of breast cancer diagnosis is significantly increased. "The accuracy, sensitivity, and specificity of the findings improved considerably, which was a good sign." The worldwide danger that breast cancer presents to women's health makes it particularly important to distinguish benign from malignant tumours based on ultrasound images. Despite the fact that both morphological and texture features are necessary for accurately representing ultrasound breast tumour images, their simple combination has little effect on the classification of benign and malignant tumours because high-dimensional texture features are too aggressive and obscure the significance of low-dimensional morphological features. An efficient textural and morphological feature combination strategy is suggested to more accurately differentiate benign and cancerous tissue. First, elements from the morphological (such as form complexity) and texture (such as local binary patterns [LBP], histograms of oriented gradients [HOG], and gray-level cooccurrence matrix [GLCM]) components of breast ultrasound images are retrieved. A support vector machine (SVM) classifier operating on texture features is trained, and a Navie Bayes (NB) classifier is constructed in order to take use of the discriminative potential of texture features and morphological traits, respectively. Thirdly, the combined weighted classification scores from the two classifiers are used to construct the final classification result (SVM and NB). The low-dimensional nonparameterized NB classifier efficiently handles the parameter complexity of the overall classification system when paired with the high-dimensional parametric SVM classifier. Thus, morphological and textural components are successfully combined. The recommended technique outperforms several comparable benign and malignant breast tumour classification methods with a 91.11% accuracy, 94.34% sensitivity, and 86.49% specificity, according to thorough experimental evaluations that are reported.

Keywords: Breast cancer detection, Machine learning classifier, Logistic regression. Performance evaluation tests Weighted sigmoid function

1 Introduction

Breast cancer is a kind of cancer that starts as a malignant tumour in the cells of the breast tissue. A malignant tumour is a collection of cancer cells that may spread across the body or infect surrounding organs [1]. Unchecked cell growth in the breast tissue is what causes breast cancer. When many cells divide fast, a lump or structural abnormalities may develop. The second most prevalent cause of death for women is breast cancer, which follows lung cancer in frequency. Early detection of breast cancer may surely reduce the mortality rate since it is a lethal disorder. An analysis of the most current data shows that the survival rate is 88% after 5 years of diagnosis and 80% after 10 years of diagnosis [2]. A significant use for machine learning classifiers is the identification of breast cancer. Numerous research efforts have been carried out in this sector. Here, a 'Logistic Regression' classifier approach has been modified to better accurately determine whether a tumorous cell is malignant or benign.

In 2015, there are anticipated to be 231,840 new instances of invasive breast cancer among women in the United States, in addition to the 60,290 new cases of non-invasive (in situ) breast cancer [3]. The leading cause of death for women in the US is breast cancer, followed by lung cancer. Breast cancer is the most prevalent kind of cancer in women globally, accounting for 25% of all cases. It affects women about 100 times more commonly than it does males, and it is more common in industrialised countries [4]. The result is influenced by the patient's age, the severity of the ailment, and the kind of breast cancer [5]. The industrialised world has high survival rates, with 80% to 90% of individuals living for at least five years in England and the United States [6, 7]. Lung cancer, which is responsible for 17% of all new instances of cancer and 23% of all cancer-related deaths, is more common in men than in women. Breast cancer is presently one of the leading causes of cancer-related mortality among women in economically developing countries [8]. The top cause of cancer-related mortality during the preceding 10 years was cervical cancer, but things have changed since then. Additionally, 11% of all female cancer deaths in underdeveloped countries are attributable to lung and cervical cancer, giving them comparable mortality rates for females. Despite the fact that overall cancer incidence rates in the developing world are for both sexes half as high as those in the developed world, overall cancer death rates are typically comparable. In Portugal, there are 4500 new instances of breast cancer identified each year, and 1600 women are anticipated to die from the condition [9].

The morphological and textural properties of breast ultrasound images are widely used to differentiate between benign and malignant tumours. The straightforward approach is to manually evaluate the texture and morphological features in images and depend on skilled and knowledgeable radiologists to distinguish between benign and malignant tumours [3]. The proportion of each component of the diagnosis in the overall judgement, however, is likely to lead to poor objectivity and repeatability of the diagnostic findings because of variations in technology and medical knowledge across physicians. Additionally, the ultrasound pictures' own high noise and low resolution significantly limit the accuracy of artificial auditory detection.

Another easy way to avoid the subjectivity of manually analysing ultrasound pictures is to train classifiers in a computer to automatically differentiate between benign and malignant tumours based on textural and morphological characteristics [4]. Computer-assisted analysis may be done in two different ways. One method for computer modelling breast pictures is to pair a single feature—one of the morphological characteristics or textural elements [5-8]—with a single classifier.

This method's insufficient consideration of the complementarity of characteristics reduces the classification's precision [5-8]. By integrating various characteristics (texture and morphological traits) with a single classifier, another method takes use of the complementarity between texture and morphological features [9–12]. Directly combining a variety of traits can reduce the efficacy of categorization, for instance, if high-dimensional texture features are overbearing and obscure the significance of low-dimensional morphological variables [13]. A single classifier cannot handle this problem. The main objective of this research is to effectively combine textural and morphological variables to improve classification performance.

For this, a technique for separating benign from malignant breast tumours is created that successfully integrates morphological and textural factors. Figure 1 depicts the proposed technique in broad strokes. It is obvious that the suggested method trains morphological and texture characteristics using two different classifiers. First, three texture features—local binary patterns (LBP), histogram of gradients (HOG), and gray-level co-occurrence matrixes (GLCM) [16]—as well as three morphological features—compactness, elliptical compactness, and radial distance spectrum—are extracted from 448 collected breast ultrasound images that have been denoised and equalised. The size of the texture features is then decreased by PAC. Second, using the support vector machine (SVM) [17] and naive bayes (NB) [18] classifiers, train texture features and morphological features. SVM already has a high-dimensional parametric classifier. Occam's razor [19] states that if one wishes to combine many classifiers, it makes logical to choose a low-dimensional nonparametric classifier to handle the parameter complexity of the overall classification system.

Thirdly, the weighted outputs of the two classifiers are combined to obtain the final classification result. Our past research projects [20, 21] have been expanded upon in this study, which improves experimental methodology. A list of the paper's contributions is provided below. (1) To improve the performance of diagnosing benign and malignant breast tumours, a novel method is proposed to more precisely incorporate several parameters and multiple classifiers. "In order to avoid the sharp increase in parameter complexity caused by using multiple classifiers, a parameterized SVM classifier trained on high-dimensional texture features is specifically designed to work in conjunction with a nonparameterized NB classifier trained on low-dimensional morphological features." (2) To illustrate the advantages of the suggested approach, extensive experimental analyses are presented, including data preprocessing, dimension reduction, single feature with single classifier, multiple features with single classifier, and successfully combining multiple features and multiple classifiers.

2 Review of literature

Computer technology has advanced significantly, and so has medical imaging technology. The practise of automatically classifying breast ultrasound images with a computer has gained popularity. This section provides an overview of breast tumour classification approaches based on hand-crafted features and deeply learned feature sets.

2.1 Hand-Crafted Features for Classification of Breast Tumours

The four processes of the standard breast tumour classification technique are image preprocessing, image segmentation, feature extraction, and tumour classification [22, 23], and they are primarily used in breast ultrasound pictures. The primary task for classifying breast tumours among them is feature extraction, which significantly affects the classification outcomes [24]. The key to analysing breast ultrasound pictures is texture (i.e., LPB [14], HOG [15], and GLCM [16]) and morphological (i.e., form complexity), also known as 'hand-crafted features.' The manually created feature-based approaches for classifying breast tumours can be loosely classified into two groups.

First, the most popular technique is to model breast ultrasound pictures with a single classifier and a single feature (one of a textural feature and a morphological feature) [5-8]. For instance, Pomponiu et al. [5] employed SVM to categorise the identified tumours after filtering tumours and normal areas based on the histogram of oriented gradients (HOG) descriptor. In order to preprocess ultrasonic pictures, Mohamed et al. [8] employed a superresolution technique and assessed the performance of five texture features.

Second, numerous techniques are used to model breast ultrasound pictures, utilising numerous features (texture features and morphological features with a single classifier) [9–12]. For instance, Menon et al. [10] employed SVM to identify cancers by extracting the textural, morphological, and histogram information from tumour ultrasound images. 41 morphological features and 96 texture features were retrieved by Gonzelezluna et al. [12] to examine the classification outcomes of 7 classifiers. Aside from SVM [17], NB [18], KNN [25], DT [26], LDA [27], and other classifiers, hand-crafted feature approaches frequently employ SVM [17], k-nearest neighbour [25], DT [26], and LDA. The two types of these classifiers are parameterized classifiers and nonparameterized classifiers. However, this type of classifier, like SVM [17] and KNN [25], has good generalisation capabilities on small data sets. In general, parameterized classifier calculation is complicated and requires repeated training to acquire the appropriate parameters. "Although the nonparameterized classifier performs poorly in terms of generalisation on tiny data sets, as noted in NB [18], it does not add any additional parameter complexity." Using two parameterized classifiers when integrating several features with distinct classifiers may overcomplicate the training model, whereas using two nonparameterized classifiers will result in weak discrimination learning [19]. In order to combine several features, a parameterized and a nonparameterized classifier are presented.

2.2 Deep-Learned Features for Classifying Breast Tumours

Deep neural networks have made dramatic strides in computer vision, thanks to improvements in computing power and very large annotated datasets [28]. The most fundamental method for classifying breast cancers using deep learning features is CNN [29]. For instance, Zhou et al. [29] and Qi et al. [30] both employed CNN to automatically identify benign and malignant tumours by extracting picture attributes. Other deeply learned characteristics are also taken into account in the classification of breast cancers. A computer-aided diagnostic system that integrates three deep learning models (Fully Convolutional Network (FCN) [32], AlexNet [33], and GoogLeNet [34]) was reviewed by Choi et al. [31] by comparing the computer's and clinicians' diagnostic outcomes.

3 Proposed methodology

3.1 Datasets and their features

There must be a dataset to categorise in order to perform classification. A statistical matrix called a dataset is used to represent various features. It is a matrix that contains all the details about the various features. The tumorous tissue feature is represented by each column in the dataset, and the number of cases is represented by each row. Three different dataset types are primarily employed in the detection of breast cancer.

- ***Image preprocessing***: The task of image preprocessing is to enhance the image and to reduce speckle without destroying the important features of BUS images for diagnosis.
- ***Image segmentation***: Image segmentation divides the image into non overlapping regions and it separates the objects (lesions) from the background. The boundaries of the lesions are delineated for feature extraction.
- ***Feature extraction and selection***: This step is to find a feature set of breast cancer lesions that can accurately distinguish between lesion and non-lesion or benign and malignant. The feature space could be very large and complex, so extracting and selecting the most effective features is very important.
- ***Classification:*** Based on the selected features, the suspicious regions will be classified into different categories, such as benign findings and malignancy. Many machine learning techniques such as linear discriminant analysis (LDA), support vector machine (SVM) and artificial neural network (ANN) have been studied for lesion classification.

Although there are certain breast ultrasound databases, accessing them is difficult to maintain patient anonymity. Since public ultrasound photos are difficult to obtain and could violate patient privacy, Quanzhou First Hospital in Fujian, China, has created a new dataset of breast ultrasound images. All of the photos were gathered between 2018 and 2019 using the 12 MHz probe frequency of the PHILIPS iu22, PHILIPS iu Elite, and other colour ultrasound diagnostic equipment. Radiologists changed the ultrasound device's imaging parameters. The relevant patients have given their permission for the usage of the photos.

After eliminating cases with missing clinical data, low picture quality, and a history of prior breast surgery, 448 breast ultrasound images could be collected. There are 264 dangerous tumours and 184 benign tumours among them.

Biopsy was performed in every case. The final classification of 448 solid breast tumours based on ultrasound findings is category 2, consider benign changes, for 43 tumours (9.6%); category 3, probably benign tumours, for 50 tumours (11.2%); category 4a, low probability of malignancy, for 91 tumours (20.3%); category 4b, median probability of malignancy, for 77 tumours (17.2%); and category 4c, high probability of malignancy, for 77 tumours (17.2%). The information gathered includes every type of tumour. the distribution of the photographs that were gathered. A highly qualified professional radiologist with more than 10 years of experience manually annotates the region of interest (ROI) and outline of tumours for each breast ultrasound image. A second expert radiologist also confirms the annotated results.

All of the photos' edges are initially erased. The computer's capacity to fully extract texture and morphological information will be constrained by the ultrasonic imaging's speckle noise and poor contrast. For this, the speckle reducing anisotropic diffusion (SRAD) filter is used to denoise every image [10]. Following that, a histogram is used to equalise the denoised photos.

3.1 Feature Extraction and Selection

As shown in Fig. 2, machine learning is used for a variety of computer jobs when it is impractical to create and construct explicit, rule-based algorithms. A training set of data comprising observations whose category membership is known allows machine learning algorithms to determine which of a set of categories a new observation belongs to. It investigates the development and research of algorithms that can learn from and forecast data. Instead, then rigidly following static programme instructions, these algorithms function by creating a model from sample inputs to produce data-driven predictions or judgements. The malignancy and benignity of tumorous cells are determined using the dataset in the classifier [22]..

Several widely used classifiers are listed below:

Naïve Bayes

• Logistic Regression

• Support Vector Machines (SVM)

• Decision Tree

• Reinforcement learning

• Neural Network

• k-Nearest Neighbors algorithm (KNN,IBK)

Here in this research work Logistic Regression classifier has been used.

1.3 Preferred use of logistic regression Logistic regression classifier is used in this research work because it implies:

- Probabilistic interpretation: The probabilistic interpretation is easy. But it is not available in Decision Tree and SVM classifier [23].
- Feature correlation: Under this mathematical model, we do not need to worry about the features being correlated unlike Naïve Bayes [24].
- Upgrade: The model can be easily updated to take in new data. This can be done using online gradient descent.
- Adjustment: Logistic regression model helps to easily adjust Classification Thresholds when we are unsure about the confidence intervals.
- Computational Considerations: Logistic regression models are usually computationally less complicated to build and require less computation time to train compared with ANNs.

3.2 Working principles of logistic regression and its features

Based on one or more predictor variables, the binary logistic regression model is used to predict binary responses (e.g. features). By estimating probabilities using a logistic function, such as the cumulative logistic distribution, logistic regression analyses the connection between the categorical dependent variable and one or more independent variables [46]. The fact that we are fitting a linear model to the feature space gives rise to the name 'regression.' A more probabilistic concept of categorization is included into logistic regression [47]. Although the result of logistic regression should not be continuous, it may be used to multidimensional feature spaces with either continuous or categorical characteristics as follows:

3.3 Texture Features.

An operation called the LBP [14] is used to represent the local texture elements of the picture and has many obvious benefits, including rotation invariance and grey invariance. A 33 window is how the LBP [14] operator is described. By comparing the size of the centre pixel value to the surrounding pixel value (often translated to LBP [14] coding, which is 256 decimal), the following 8-bit binary number is created:

$$\mathrm{LBP}(x, y) = \sum_{p=1}^{8} 2^{p-1} s\left(i_p - i_c\right), \qquad (1)$$

$$s(x) = \begin{cases} 1 \text{ if } x \geq 0 \\ 0 \text{ else} \end{cases},$$

where i is the gray value of the center pixel ðx, yÞ, p is the number of the adjacent pixel, ip is the gray value of the adjacent pixel, and sðxÞ is the symbolic function.

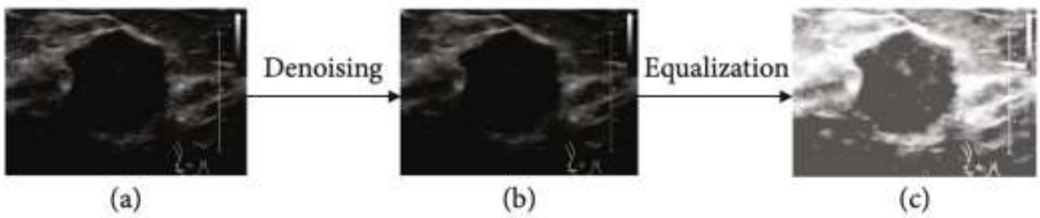

Fig 1: The result after using SRAD filter and histogram to denoise and equalize the breast ultrasound images: (a) shows the original image, (b) shows the denoised image, and (c) shows the result after equalization. In order to create the feature, the HOG [15] first computes and statistically analyses the histogram of gradient direction in the immediate picture region. "Each pixel's gradient is determined after Gamma correction and the picture is then shown." Second, this work uses a grid of 32 by 32-pixel cells to analyse images and generate a descriptor based on the histogram of gradient inside each cell. The HOG [15] feature descriptor is obtained by first stringing together each 22 cell into a block, and then stringing together all blocks. The GLCM [16] is used to determine the connection between each pair of pixels. A grayscale value of 64 was used for this publication. We change the spacing between pixels anywhere from [1, 10], and we compute the connection between pairs of pixels based on their distance from one another in four orthogonal directions (0, 45, 90, 135). At the end of the process, 40 unique matrices are extracted from each picture. "Matrices are mined for energy, contrast, correlation, and homogeneity, all of which are indicative of the image's textural roughness, local variation, and overall uniformity of its grey distribution."

3.4 Morphological Features.

Compaction (equation 2), elliptic compactness (equation 3), and the mean and variance of the radial distance spectrum (equation 4) are all useful in determining the tumor's morphology. If the tumour has an uneven lobule form rather than being perfectly round or oval, it may be malignant [8]. A breast tumor's compactness indicates how closely its form matches that of the centre circle. A tumour is less likely to be malignant if its compactness value is near to 1.

$$C = \frac{A}{4\pi L^2}, \qquad (2)$$

where A is the tumour size in square metres and L is the length of the breast tumour outline in centimetres. The elliptic compactness measures how well the ellipse fits the original tumour outline in comparison to its own diameter. It's inversely connected to how aggressive a tumour is. Finding an ellipse that best fits a collection of tumour contour points is the goal of the elliptic fitting technique. In a broader sense, the elliptic equation is used as a model to fit the tumor's contour points in order to find an elliptic equation that can fulfil these points as well as feasible, and the values of all the elliptic equation's parameters are then determined. In this case, we fit an ellipse using the least squares approach, as suggested by Fitzgibbon et al. [36]. Figure 5 depicts the outcome of ellipse fitting. The tumor's border (blue) and the corresponding ellipse's (red) outline are shown below. The following characteristics are computed based on the ellipse of best fit obtained:

$$EC = \frac{\pi(a+b)}{D}, \qquad (3)$$

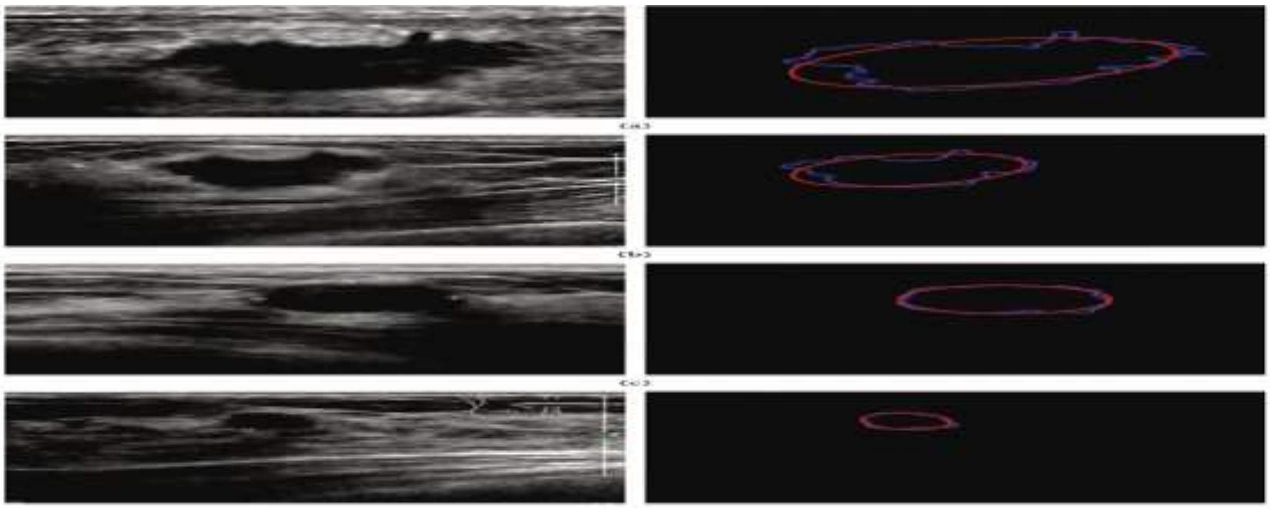

Fig 2 : The examples of the fitting ellipse that transformed from breast tumor contour: (a, b) malignant tumor and (c, d) benign tumor. where an is the fitting ellipse's semimajor axis, b is the semiminor axis, and D is the breast tumour contour's perimeter. By statistically and analytically assessing the radial distance from each location on the tumour margin to the tumour centre, the radial distance spectrum approach was able to quantify the degree of tumour margin roughness. In this study, we apply Fourier transform on the radial distance spectrum and then take the logarithm of the resulting spectrum to produce the radial distance spectrum with logarithmic amplitude. The logarithmic amplitude spectrum is then analysed for its harmonic components, and its mean and variance are used as defining statistics. For a radial distance, use these formulas:

$$D(t) = \sqrt{(p_t - x_0)^2 + (q_t - y_0)^2}, \qquad (4)$$

where the tumor edge points are denoted as Ptðpt, qtÞ and the center point is denoted as ðx0, y0Þ.

Feature Selection. In this paper, the principal component analysis (PCA) [37] is used to reduce the dimension of extracted texture features in order to speed up the training and testing time and improve the efficiency of the proposed method.

4 Experiment and result

4.1 Experimental Results.

In the following, a comparison is used to confirm the outcome of the suggested procedure. In this paper, texture features (local binary patterns (LBP) [14], histogram of gradients (HOG) [15], and gray-level co-occurrence matrices (GLCM) [16]) and morphological features (compactness, elliptical compactness, and radial distance spectrum) [17]) are learned efficiently using support vector machine (SVM) [17] and naive Bayes (NB) [18] classifiers, respectively. This study contrasts the suggested approach with similar approaches [5, 7, 8, 10, 12] to demonstrate its superiority. Matlab 2017b is mostly used to conduct the studies. Table 1 shows that the hand-crafted feature technique is able to learn from a limited sample and improve classification accuracy. Furthermore, experimental findings demonstrate that the classification performance of multiple features is typically superior than single feature, and our approach makes full use of the complementarity of texture and morphological characteristics to acquire the higher performance than single classifier. Our strategy outperforms similar approaches, with a 91.11% accuracy rate, 94.34% sensitivity, and 86.49% specificity. Breast tumours may be classified as benign or malignant depending on whether or not many characteristics and classifiers are used.

4.2 Discussion.

The following analysis and discussion is carried out to demonstrate the efficacy of the suggested strategy. In addition to SVM [17] and NB [18], three classifiers (k-nearest neighbour (KNN) [25], decision tree (DT) [26], and linear discriminant analysis (LDA) [27]) are used to examine the three retrieved texture characteristics (LBP [14], HOG [15], and GLCM [16]). (com- pactness, elliptical compactness, and radial distance spec- trum). The next three parts outline the methodology that will be used to conduct the experimental analysis.

Table 1: The performance comparison of our method and multiple related methods.

Method		Evaluation (%) Accuracy	Sensitivity	Specificity
Single feature with single classifier (SFSC)	Pomponiu et al. [5]	81.11	84.91	75.68
	Biswas et al. [7]	75.56	67.92	86.49
	Mohamed et al. [8]	84.44	84.91	83.78
Multiple features with single classifier (MFSC)	Menon et al. [10]	87.78	88.68	86.49
	Gonzelezluna et al. [12]	86.67	88.68	83.78
Multiple features with multiple classifiers (MFMC)	Our method	91.11	94.34	86.49

Table 2: The classification results based on the methods of single features with single classifier.

Method Feature	Classifier	Accuracy	Evaluation (%)Sensitivity	Specificity
LBP	SVM [17]	85.56	86.79	83.78
	KNN [25]	84.44	84.91	83.78
	DT [26]	66.33	58.49	81.08
	LDA [27]	74.44	77.36	70.27
HOG	SVM [17]	81.11	84.91	75.68
	KNN [25]	61.11	100.00	5.41
	DT [26]	67.78	67.92	67.57
	LDA [27]	70.00	75.47	62.16
GLCM	SVM [17]	78.89	92.45	59.46
	KNN [25]	65.56	75.47	51.35
	DT [26]	71.11	77.36	62.16
	LDA [27]	74.44	84.91	59.46
LBP+HOG+GLCM	SVM [17]	86.67	92.45	78.38
	KNN [25]	64.44	100.00	13.51
	DT [26]	72.22	73.58	70.27
	LDA [27]	75.56	84.91	62.16
Morphological	SVM [17]	75.56	67.92	86.49
	NB [18]	81.11	69.81	97.30
	LDA [26]	75.56	60.38	97.30

4.2.1 Experiments Based on Classification Methods Using Single Features with Single Classifier

Table 2 shows the specifics of the sensitivity, specificity, and accuracy of the model prediction based on LBP [14], HOG [15], GLCM [16] texture features, fused texture features, and morphological characteristics. "The classification results based on the fused texture features are the best when compared to the classification results of the various texture characteristics in Table 2, with an accuracy of 86.67%, a sensitivity of 92.45%, and a specificity of 78.38%. LBP [14] is the second-best characteristic, with an accuracy of 85.56%, sensitivity of 86.79%, and specificity of 83.78%." HOG [15] feature has a high rate of accuracy (81.11%), sensitivity (84.91%), and specificity (75.68%). GLCM [16] has a 78.89% accuracy, a 92.45% sensitivity, and a 59.46% specificity. There is an 81.11 percent precision, a 69.81 percent sensitivity, and a 97.3 percent specificity when analysing morphological characteristics.

Table 2 shows the results of many different classifiers, and it can be seen that the SVM [17] classifier achieves the highest classification results overall, with an accuracy of 86.67 percent, a sensitivity of 92.4 percent, and a specificity of 78.5 percent. The KNN [25] classifier comes in second with a sensitivity of 84.91%, specificity of 83.08%, and accuracy of 84.44%. In terms of accuracy, the NB [18] classifier scores 81.11 percent; in terms of sensitivity, it scores 69.81 percent; and in terms of specificity, it scores 97.3 percent. LDA [27] classifier has an accuracy of 75.56%, sensitivity of 60.38%, and specificity of 97.30%. It has been shown that the DT [26] classifier can achieve an accuracy of 72.22 percent, a sensitivity of 73.58 percent, and a specificity of 60.27 percent.

Comparing the experimental findings in Table 2, we find that the fused texture features (LBP+HOG+GLCM) with the SVM classifier achieve the best results among the single feature and single classifier classifica- tion techniques. These approaches use a single feature with a single classifier, but the complementarity of features is not completely explored, which limits the accuracy of classification.

4.2.1 Experiments Based on Classification Methods Using Multiple Features with Single Classifier

Table 2 shows the specifics of the sensitivity, specificity, and accuracy of the model prediction based on LBP [14], HOG [15], GLCM [16] texture features, fused texture features, and morphological characteristics. The accuracy of the classification results based on the fused texture features reaches 86.67%, the sensitivity reaches 92.45%, and the specificity reaches 78.38%, which is the best of all the classification results of the various texture features in Table 2. LBP [14], the second-best feature, is accurate 85.56% of the time, sensitive 86.79% of the time, and specific 83.78% of the time. HOG [15] has an 81.11 percent precision, an 84.91 percent sensitivity, and a 75.68% specificity. GLCM [16] has a 78.89% accuracy, a sensitivity of 92.45%, and a specificity of 59.56%. The accuracies, sensitivities, and specificities for morphological characteristics are 81.11, 69.81, and 97.30 percent, respectively. From Table 2, it is clear that the SVM [17] classifier achieves the highest overall classification results, with an accuracy of 86.67%, a sensitivity of 92.45%, and a specificity of 78.38%. This is in contrast to the classification results of the other classifiers. KNN [25] is the second classifier, and it has an accuracy of 84.44%, sensitivity of 84.91%, and specificity of 83.08%. For its part, the NB [18] classifier is 81.11 percent accurate, 69.81 percent sensitive, and 97.3 percent specific. The LDA [27] classifier has an accuracy of 75.56%, sensitivity of 60.38%, and specificity of 97.30%. "In terms of accuracy, DT [26] classifier can achieve 72.22 percent, with a sensitivity of 73.58 percent and a specificity of 60.27 percent. According to Table 2's comparison of experimental findings, the fused texture features (LBP+HOG +GLCM) with SVM classifier achieve the best results among the single feature and single classifier classification techniques." However, the accuracy of classification is limited by these approaches since they only use a single feature with a single classifier, without taking into account the complementarity of features.

TABLE 3: The classification results based on the methods of multiple features with single classifier.

Features	Evaluation (%) Classifier	SVM	LDA
LBP+ morphological	Accuracy	80.00	76.67
	Sensitivity	79.25	75.47
	Specificity	81.08	78.38
HOG+ morphological	Accuracy	83.33	72.22
	Sensitivity	81.13	71.70
	Specificity	86.48	72.97
GLCM+ morphological	Accuracy	80.00	80.00
	Sensitivity	69.81	81.13
	Specificity	94.59	78.38
LBP+HOG+GLCM+ morphological	Accuracy	87.78	76.67
	Sensitivity	88.68	75.47
	Specificity	86.49	78.38

TABlE 4: The classification results based on the method of multiplefeatures with multiple classifiers.

Method	Evaluation (%)		
	Accuracy	Sensitivity	Specificity
SVM (LBP+HOG+GLCM) [17]	86.67	92.45	78.38
NB (morphological) [18]	81.11	69.81	97.30
Our method	91.11	94.34	86.49

4.2.2 Effect Analysis of Image Pre-processing and Feature Selection.

Categorization experiments are conducted with and without denoising and equalisation applied to ultrasound pictures, respectively, to establish that these processes have a supporting role in the classification of ultrasound images. This research analyses studies that simply extract texture characteristics since noise and contrast in the picture have minimal impact on the morphological traits. The texture characteristics are analysed beforehand using the SVM classifier with the highest classification performance. Table 5 demonstrates how pretreatment enhances the precision with which texture characteristics are extracted from photos. First-time texture feature extraction has a too-high dimensionality. To keep only the most useful features, PCA can be used to shrink the eigenvector matrix. As can be observed in Table 6, the testing time needed for the classification approach as a whole is considerably decreased following dimension reduction of texture characteristics. By reducing the number of dimensions, training time may be cut down to 0.0135 seconds, or 1/46 of the original time.

5 CONCLUSION

In order to enhance ultrasonic imaging's classification performance for benign and malignant tumours, this research proposes an approach that combines textural and morphological features. First, 448 breast tumour ultrasound images are denoised and equalised before texture features (i.e. local binary patterns (LBP), histogram of oriented gradients (HOG), and gray-level co-occurrence matrixes (GLCM)) and morphological features (i.e. compactness, elliptical compactness, and radial distance spectrum) are extracted. Second, as high-dimensional texture characteristics may readily impact low-dimensional morphological data in a single classifier, we utilise separate classifiers to learn texture features and morphological features. When both classifiers have finished their work, their scores are combined using a weighted fusion to get a final classification. Using the high-dimensional parametric SVM classifier in conjunction with the low-dimensional nonparameterized NB classifier allows for effective management of the parameter complexity of the overall classification system. Extensive experimental analyses are presented to validate the efficacy of the proposed method, in which a total of three classifiers (i.e., k-nearest neighbour (KNN), decision tree (DT), and linear discriminant analysis [LDA]) are used to analyse the three extracted texture features and three morphological features. Single-feature and multi-feature analyses are among the options. "Based on experimental findings, it is clear that the suggested strategy is the most effective in terms of accuracy, sensitivity, and specificity." This is a feasible and robust way to examine breast cancers using ultrasound in a short amount of time and at a reasonable cost.

REFERENCES

i. F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and "A. Jemal,Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: A Cancer Journal for Clinicians, vol. 68, no. 6, pp. 394–424, 2018.

ii. C. E. DeSantis, J. Ma, A. Goding Sauer, L. A. Newman, and A. Jemal, 'Breast cancer statistics, 2017, racial disparity in mortality by state,' CA: A Cancer Journal for Clinicians, vol. 67, no. 6, pp. 439–448, 2017.

iii. T. Steifer and M. Lewandowski, 'Ultrasound tissue characterization based on the Lempel–Ziv complexity with application to breast lesion classification,' Biomedical Signal Processing and Control, vol. 51, pp. 235–242, 2019.

iv. H. D. Cheng, J. Shan, W. Ju, Y. Guo, and L. Zhang, 'Automated breast cancer detection and classification using ultrasound images: a survey,' Pattern Recognition, vol. 43, no. 1, pp. 299–317, 2010.

v. V. Pomponiu, H. Hariharan, B. Zheng, and D. Gur, 'Improving breast mass detection using histogram of oriented gradients,' Medical Imaging 2014: Computer-Aided Diagnosis, vol. 9035, article 90351R, 2014International Society for Optics and Photonics. [6] A. A. Ardakani, A. Gharbali, and A. Mohammadi, 'Classification of breast tumors using sonographic texture analysis,' Journal of Ultrasound in Medicine, vol. 34, no. 2, pp. 225– 231, 2015.

vi. R. Biswas, A. Nath, and S. Roy, 'Mammogram classification using gray-level co-occurrence matrix for diagnosis of breast cancer,' in 2016 International Conference on Micro-Electronics and Telecommunication Engineering (ICMETE), pp. 161–166, Ghaziabad, India, 2016.

vii. M. Abdel-Nasser, J. Melendez, A. Moreno, O. A. Omer, and D. Puig, 'Breast tumor classification in ultrasound images using texture analysis and super-resolution methods,' Engineering Applications of Artificial Intelligence, vol. 59, pp. 84–92, 2017.

viii. W. Gómez Flores, W. C. . A. Pereira, and A. F. C. Infantosi, 'Improving classification performance of breast lesions on ultrasonography,' Pattern Recognition, vol. 48, no. 4, pp. 1125–1136, 2015.

ix. R. V. Menon, P. Raha, S. Kothari, S. Chakraborty, I. Chakrabarti, and R. Karim, 'Automated detection and classification of mass from breast ultrasound images,' in 2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), pp. 1–4, Patna, India, 2015.

x. A. Rodríguez-Cristerna, W. Gómez-Flores, and W. C. de Albuquerque Pereira, 'A computer-aided diagnosis system for breast ultrasound based on weighted BI-RADS classes,' Computer Methods and Programs in Biomedicine, vol. 153, pp. 33–40, 2018. [12] F. A. González-Luna, J. Hernández-López, and W. GomezFlores, 'A performance evaluation of machine learning techniques for breast ultrasound classification,' in 2019 16th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), pp. 1–5, Mexico City, Mexico, 2019.

xi. W. C. Shen, R. F. Chang, W. K. Moon, Y. H. Chou, and C. S. Huang, 'Breast ultrasound computer-aided diagnosis using BI-RADS features,' Academic Radiology, vol. 14, no. 8, pp. 928–939, 2007.

xii. T. Ojala, M. Pietikäinen, and D. Harwood, 'A comparative study of texture measures with classification based on featured distributions,' *Pattern Recognition*, vol. 29, no. 1, pp. 51–59,1996.

xiii. N. Dalal and B. Triggs, 'Histograms of oriented gradients for human detection,' in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol.1, pp. 886–893, San Diego, CA, USA, 2005.

xiv. R. M. Haralick, K. Shanmugam, and I. H. Dinstein, 'Textural features for image classification,' *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.

xv. C. C. Chang and C. J. Lin, 'LIBSVM: A library for support vec- tor machines,' *ACM transactions on intelligent systems and technology (TIST)*, vol. 2, no. 3, pp. 1–27, 2011.

xvi. Rish, 'An empirical study of the naive Bayes classifier,' *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, vol. 3, pp. 41–46, 2001.

xvii. V. Balasubramanian, 'Statistical inference, Occam's razor, and statistical mechanics on the space of probability distributions,' *Neural Computation*, vol. 9, no. 2, pp. 349–368, 1997.

xviii. M. Wei, X. Wu, J. Zhu et al., 'Multi-feature fusion for ultra- sound breast image classification of benign and malignant,' in *2019 IEEE 4th International Conference on Image, Visionand Computing (ICIVC)*, pp. 474–478, Xiamen, China, 2019.

xix. M. Wei, Y. Du, X. Wu, and J. Zhu, 'Automatic classification of benign and malignant breast tumors in ultrasound image with texture and morphological features,' in *2019 IEEE 13th Inter- national Conference on Anti-counterfeiting, Security, and Identification (ASID)*, pp. 126–130, Xiamen, China, 2019.

xx. K. M. Meiburger, U. R. Acharya, and F. Molinari, 'Automated localization and segmentation techniques for B-mode ultrasound images: a review,' *Computers in Biology and Medicine*, vol. 92, pp. 210–235, 2018.

xxi. M. Elawady, I. Sadek, A. E. R. Shabayek, G. Pons, and S. Ganau, 'Automatic nonlinear filtering and segmentation for breast ultrasound images,' in *International Conference on Image Analysis and Recognition*, pp. 206–213, Póvoa deVarzim, Portugal, 2016.

xxii. W. K. Moon, I. L. Chen, J. M. Chang, S. U. Shin, C. M. Lo, and R. F. Chang, 'The adaptive computer-aided diagnosis system based on tumor sizes for the classification of breast tumors detected at screening ultrasound,' *Ultrasonics*, vol. 76, pp. 70–77, 2017.

xxiii. R. Min, D. A. Stanley, Z. Yuan, A. Bonner, and Z. Zhang, 'A deep non-linear feature mapping for large-margin knn classification,' in *2009 Ninth IEEE International Conference on Data Mining*, pp. 357–366, Miami, FL, USA, 2009.

xxiv. S. R. Safavian and D. Landgrebe, 'A survey of decision tree classifier methodology,' *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.

xxv. L. J. Hargrove, E. J. Scheme, K. B. Englehart, and B. S. Hudgins, 'Multiple binary classifications via linear discriminant analysis for improved controllability of a powered prosthesis,' *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, no. 1, pp. 49–57, 2010.

xxvi. Q. Huang, F. Zhang, and X. Li, 'Machine learning in ultra- sound computer-aided diagnostic systems: a survey,' *BioMed Research International*, vol. 2018, Article ID 5137904,10 pages, 2018.

xxvii. Y. Zhou, J. Xu, Q. Liu et al., "A radiomics approach with CNNfor shear-wave elastography breast tumor classification," *IEEE*