

Article

Mammography Image-Based Diagnosis of Breast Cancer Using Machine Learning: A Pilot Study

Maha M. Alshammari ^{1,*}, Afnan Almuhanha ² and Jamal Alhiyafi ³

¹ Computational Unit, Department of Environmental Health, Institute for Research and Medical Consultations, Imam Abdulrahman Bin Faisal University, Dammam 31441, Saudi Arabia

² Department of Radiology, College of Medicine, Imam Abdulrahman Bin Faisal University, Dammam 31441, Saudi Arabia; amuhanna@iau.edu.sa

³ Department of Computer Science, Kettering University, Flint, MI 48504, USA; jalhiyafi@kettering.edu

* Correspondence: mmashammari@iau.edu.sa

Abstract: A tumor is an abnormal tissue classified as either benign or malignant. A breast tumor is one of the most common tumors in women. Radiologists use mammograms to identify a breast tumor and classify it, which is a time-consuming process and prone to error due to the complexity of the tumor. In this study, we applied machine learning-based techniques to assist the radiologist in reading mammogram images and classifying the tumor in a very reasonable time interval. We extracted several features from the region of interest in the mammogram, which the radiologist manually annotated. These features are incorporated into a classification engine to train and build the proposed structure classification models. We used a dataset that was not previously seen in the model to evaluate the accuracy of the proposed system following the standard model evaluation schemes. Accordingly, this study found that various factors could affect the performance, which we avoided after experimenting all the possible ways. This study finally recommends using the optimized Support Vector Machine or Naïve Bayes, which produced 100% accuracy after integrating the feature selection and hyper-parameter optimization schemes.

Keywords: breast cancer; machine learning; classification; support vector machine; decision tree; K-nearest neighbor; Naïve Bayes; discriminant analysis; benign; malignant



Citation: Alshammari, M.M.; Almuhanha, A.; Alhiyafi, J. Mammography Image-Based Diagnosis of Breast Cancer Using Machine Learning: A Pilot Study. *Sensors* **2022**, *22*, 203. <https://doi.org/10.3390/s22010203>

Academic Editor: Manuchehr Soleimani

Received: 6 December 2021
Accepted: 24 December 2021
Published: 28 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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

Breast cancer was ranked the fifth among the most common causes of cancer death based on World Health Organization (WHO) statistics [1]. Breast cancer is considered a frequent cancer, where 2.1 million women worldwide get affected every year [2]. The estimated mortality rate of this cause in 2020 was 685,000, which was approximately 13.6% of all cancer deaths that involved women [2]. In the Kingdom of Saudi Arabia, it represents about 15.9% of the detected cases of cancer, according to the latest Ministry of Health statistics in 2014 [3]. It is ranked high among various types of cancer that affect Saudi citizens [3]. Therefore, early detection of breast cancer is essential to reduce its mortality rate and increase the recovery rate [2,4]. Most diagnosed women are detected in the late stages due to resource limitations [2]. The screening technique is one of the earliest detection strategies before seeing the breast cancer symptoms. Many tools are used as screening techniques, such as mammography, which is considered the most common tool for early detection of breast cancer [2,4]. Mammography images produce a significant amount of information that a radiologist has to analyze and assess comprehensively in a short time to detect the type of tumor in its early stage, while the multistage process usually takes a longer time.

This paper proposes a machine learning (ML)-based computer-aided diagnosis (CAD) system to assist radiologists in classifying breast tumors as either benign or malignant using mammography data. It can rank the type of tumor without going through multiple

stages that may cause treatment delay. To accomplish the aim of this paper, we propose a new method for developing a CAD system for breast cancer. The main idea of this method is to integrate the radiologist in the tumor annotation step in the CAD's workflow since an efficient CAD system should not completely exclude the radiologist's role [5]. We extracted several features using the annotated tumor to train the classifier for classification purposes. We tested the performance of five classifiers using several experiments. These classifiers are Decision Tree (DT), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Naïve Bayes (NB), and Discriminant Analysis (DA).

We selected the classifiers based on reasonable grounds. We chose them to represent the different models' schemes, such as statistical-based, divide and conquer-based, instance-based learning, and linear classification-based [6].

In addition, we selected each classifier based on its advantages. DT is based on a divide and conquer (rules-based) scheme [6]. It is one of the popular approaches for classification [7], and it is easy to understand due to its mechanism [8]. Learning processes based on instance are represented by the KNN technique [6], which is a more straightforward technique that identifies the group's membership using similarity [9,10], especially with the numeric values [6]. NB is a probabilistic-based scheme [6], where it works according to a simple approach [10,11] that uses probabilistic theory [6,12]. It works effectively in practice with excellent results for classification [6,13]. It performed well with breast cancer problems in previous studies [10]. Based on the linear classification principles, SVM and DA were selected. Mathematically, SVM has a strong basis in theory. It showed very high performance in various disease diagnoses, image processing, etc. [13]. On the other hand, the discriminant function works well for group separation purposes, making it significant [14].

The proposed ML-based CAD system can assist the radiologist in Saudi Arabia in interpreting such images and thus improve the performance of diagnosing breast cancer and support faster decision-making for the treatment.

We used MATLAB in this work. Three experiments were done on the above classification techniques using different approaches to improve the performance. The process started after extracting 12 features based on the breast density, shape, and texture using a local real mammography dataset that never has been used previously. Each experiment consisted of several tests using different criteria. Accordingly, this paper found that various factors can affect performance. After testing all possible ways, these factors were avoided in the third experiment, clarifying those factors. After integrating the feature selection and hyper-parameter optimization functions, we obtained excellent accuracy for all tested classifiers. We did a proving test using unlabeled samples, and a radiologist reviewed the results and indicated that the optimized SVM and NB obtained 100% accuracy.

The remaining part of this paper is organized as follows: Section 2 reviews several related works; Section 3 explains the proposed method; Section 4 contains the experimental studies and results; Section 5 discusses the findings of the proposed method and the recommended models; and Section 6 concludes this paper.

2. Background

CAD systems play an essential role in the preliminary stages of detecting disease. Different techniques of ML can be used individually or in hybrid. The technique selection depends on the available data, its ambiguity, and the ability to select useful features [15,16].

One of the most critical CAD systems is the system that aims to detect or diagnose breast cancer utilizing computer technology to detect any abnormalities in breast tissues [17]. The systems are classified as one of three systems: abnormality-detection systems classify the abnormality to either benign or malignant or detect and classify the class of abnormalities [18]. The CAD system is an essential alternative to the biopsy due to the possible effects of biopsy, such as discomfort to the patient, infection, bleeding, and the time needed to get the results after laboratory analysis [19].

Significant early stage detection of breast cancer relies on the breast imaging methods used. The accuracy of interpretation by radiologists of the mammograms depends on

different factors. Mammograms have been included in CAD systems [20], which is one of the most critical advances in CAD systems to serve this purpose [18,20]. The typical stages in mammography-based CAD systems are shown in Figure 1 [18].



Figure 1. Mammography-based CAD system.

Many classification techniques can be examined associated with the extracted features to evaluate the performance of each of them. The most popular classification techniques that we reviewed are the following.

2.1. Decision Trees (DT)

A Decision Tree is divided based on the rules [6]. It is a structural method represented as a flowchart similar to the tree and more like the human's logic, making it easy to understand [8].

The critical characteristic of a Decision Tree is how it is easy to read by people due to its mechanism [8]. It combines categorical and numerical data [6,7]. It has a few issues and challenges, such as classification problems with high dimensionalities and unbalanced data [8].

2.2. K-Nearest Neighboring (KNN)

KNN is an instance-based classifier [6]. It is an efficient prediction model and one of the most straightforward techniques for learning artificial algorithms based on similarity [9,10,21]. The most affected factor of KNN, especially in medical applications, is K's value, and there are different ways of selecting its value. The easiest way could be done by running the KNN algorithm many times and changing the K value every time until the optimal value is found [21].

It performs very well in some cases. It has a few limitations, such as the time it takes to check all instances with each instance classification; also, its sensitivity to any type of attribute, correlated or not, and any mistake in specifying the value of k can affect its performance [9].

2.3. Support Vector Machine (SVM)

SVM is statistical learning-based model that works according to the supervised learning approach [13,22–24]. It belongs to the linear classification models, which works with low-dimensional and uncomplicated data that can easily separate the data space. However, it can also deal with a non-linear classifications that has high-dimensional and complex data by using a kernel function to transform the data into another high-dimensional data space that can be separated [13,22–25]. It represents the data items as points and classifies them with their belonging classes in the space. It works by finding the hyperplane that separates them by maximizing the margins that differentiate between the classes [22,23,26–28].

It is a powerful and robust model; its efficiency has been proven in various applications, such as financial, biological, image processing, bioinformatics, face recognition, text mining, and disease diagnosis for different medical fields, especially in cancer prediction and classification. The SVM efficiency in the applications mentioned above makes it a prevalent technique [13,29–31]. Mathematically, it has a strong basis in statistical learning theory [13].

2.4. Naïve Bayes (NB)

Naïve Bayes is a simple supervised learning technique for classification based on the probabilistic theorem [6,10,11]. It is a very efficient technique, especially when combined with feature selection functions [6].

It outperforms other powerful techniques due to many reasons, such as being simple to apply, fast, and producing high accuracy. It has been used in many areas, such as disease diagnosis, classification of RNA sequences, and spam filtering [32].

2.5. Discriminant Analysis (DA)

Discriminant Analysis is appropriate to be used in supervised learning for data grouping tasks that have known classes. It works by predefined groups using original data to classify new unlabeled instances [33,34]. It is a statistically based technique and one of the linear classifiers [14].

This technique is multivariate; it works by finding the best combinations of the attributes used as predictors to maximize the separation between the given groups of data, which is needed to predict the group membership of any new instance. The best combinations between the given attributes are called discriminant functions [14].

3. Related Work

Recently, many researchers have highlighted intelligent systems. Many disease diagnoses and decision support have included studies to diagnose tumors, especially breast tumors. Some studies focused on using a single technique, and others used a combination of techniques to achieve the highest reliability of the results depending on different data.

This section reviewed the state-of-the-art studies in breast cancer diagnosis using machine learning techniques from 2007 to the present time. The review process started in 2007 due to the gap in studies done before 2007. In 2007, this subject became very fertile and an interesting research subject. Awareness of early detection of breast cancer increased in the world. There is a need to employ artificial intelligence applications to support this process and help radiologists support their treatment decisions.

In 2007, Polat et al. [35] proposed a hybrid method, where he combined a fuzzy-artificial immune system with KNN algorithms. He classified one of the most common datasets, Wisconsin Diagnostic Breast Cancer (WBCD), to examine the proposed method's accuracy compared to other studies. The proposed technique achieved 99.14% accuracy, and its accuracy was obtained using one of the cross-validation techniques.

Akay [36] proposed the SVM-based model, and it was combined with feature selection. The model was trained and evaluated using different partitions of WBCD. The SVM model achieved 99.51% accuracy.

Keles et al. [19] developed expert system-based neuro-fuzzy rules for diagnosing breast cancer (Ex-DBC) to avoid using biopsy, and it used an excellent performance learning algorithm in disease diagnosis and classification. The proposed system contained three layers: the input layer, the hidden layer containing nine fuzzy-rules, and the last layer representing the output that specified the class of a particular case.

A new Neuro-Fuzzy System (NFS) [37] has been proposed and its performance compared with two other powerful supervised techniques: Radial Basis Function Neural Network (RBFNN) and Adaptive Neuro-Fuzzy Inference System (ANFIS). The training and testing phases for the RBFNN, ANFIS, and NFS techniques were done using the WBCD dataset. NFS achieved the highest classification accuracy, which was 98.4%.

A digital mammography dataset was used to evaluate the new algorithm [38]. The proposed algorithm was based on Non-Subsampled Contourlet Transform (NSCT) and Super Resolution (SR) combined with an AdaBoost classifier. NSCT and SR improved the image quality after automating the extraction of the region of interest (ROI). An accuracy of 91.43% was achieved.

A novel CAD system has been proposed [39] based on the Extreme Learning Machine (ELM) classifier using the available mammography images set. It included several steps: image preprocessing, ROI segmentation, features extraction, features selection, mass classification, and final assessment of the proposed model. The proposed ELM obtained 96.02% accuracy.

Emina and Abdulhamit [40] proposed a two-stage system. First, they used a genetic algorithm to extract the essential and useful features from a dataset containing 569 cases and 32 features. In the second stage, hybrid and individual data mining techniques were used to propose an accurate result. The Rotation Forest technique obtained the highest accuracy, which was 99.48%.

Hybridized SVM algorithms based on Weighted Area Under the Receiver Operating Characteristic Curve Ensemble (WAUCE) were incorporated to build a new model to increase the accuracy and reduce a variety of breast cancer diagnoses [41]. Two published and one real dataset were used to evaluate the proposed model. The highest accuracy achieved was 97.68% by applying a 10-fold validation technique on the Wisconsin Diagnostic Breast Cancer (WDBC) dataset.

Using the 569 cases and 32 features of the dataset, Joshi and Mehta [42] examined the diagnosis accuracy of the KNN ML technique in the R environment. It was done by employing the most significant dimensionality reduction techniques: Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA). LDA achieved a better accuracy (97.06%) than PCA.

A novel intelligent model has been proposed. It is based on the combination of an information gain algorithm for features ranking and a simulated annealing genetic algorithm wrapper for selecting the optimal features [43]. The chosen features inferred the cost-sensitive SVM classifier. The Wisconsin Original Breast Cancer (WBC) and WDBC datasets were used to train and test the proposed model. The optimum accuracy was 95.8%, which was obtained with the WBC dataset.

Both Chaurasia and Pal [44] and Abdar et al. [45] used the WDBC dataset to compare different ML techniques' accuracy. In Abdar et al.'s study [45], an approach of the nested ensemble was applied by using two layers. It contained two to three classifiers that have been combined using Stacking and Vote (SV); the SV-NB-3-Meta Classifier had the best obtained accuracy (98.07%) with the least model building time compared with the other combined algorithms. Chaurasia and Pal [44] compared six ML techniques, including SVM, NB, KNN, DT, and Random Forest (RF); all of them exceeded 90% accuracy.

4. Proposed Method

A series of image processing algorithms were developed to build the proposed ML-based CAD system containing multiple stages. The preliminary stages shown in Figure 2 are discussed in the following subsections.

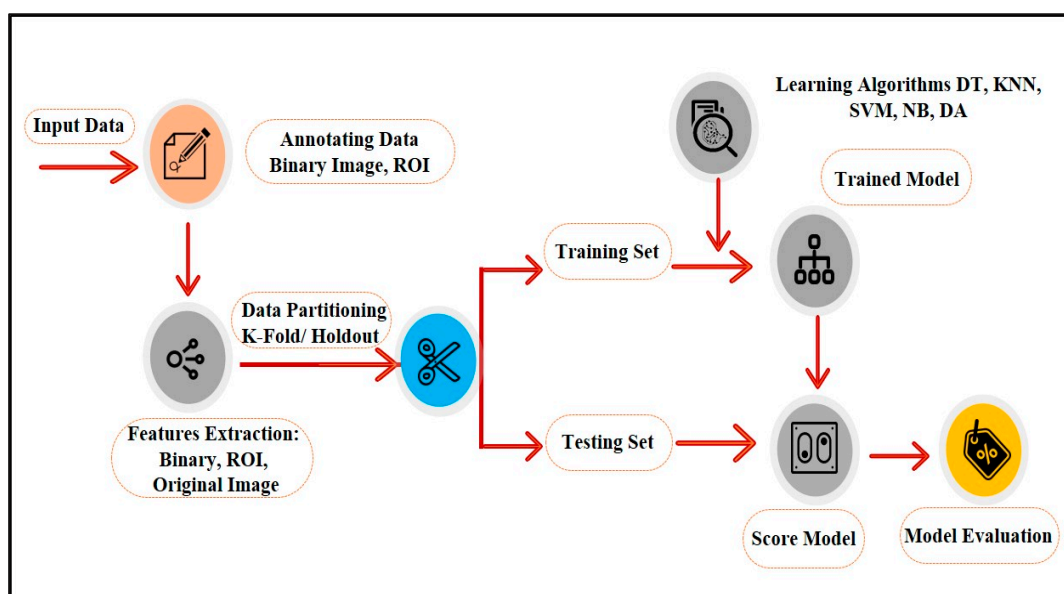


Figure 2. Schematic diagram for the proposed ML-based CAD system.

4.1. Annotating Data

One of the reasonable approaches to describe the abnormalities is by human perception. While the radiologists have special training to identify the state of the tissue in the breast from its appearance in a mammogram, many CAD systems used the features described by radiologists to build and train the proposed systems, although they used different terminologies for the same features [46].

Other CAD systems for breast cancer diagnosis based on mammography data used image processing techniques for segmenting and extracting features, which may affect the accuracy due to the tissue ambiguity and complexity.

In this study, the nominated radiologist manually annotated the abnormal region by drawing the border between the tumor and the background of a mammogram.

4.2. Features Extraction

In this stage, algorithms were developed to extract several features to characterize the ROI that the radiologist annotated. These features are intensity-based (average intensity, standard deviation, and contrast between the foreground ROI and background), shape-based (degree of circularity, elongation, length, and diameter), or texture-based features. Texture is a significant property to identify ROI [47,48], and they provide characteristics such as smoothness, coarseness, and regularity of the intensity values in a ROI [26,49].

The main problems in texture features analysis are specifying which features are required and the processing type needed for texture recognition [50]. The effective texture features depend on the type of images and effective factors in the research problem. Texture features extraction requires an evaluation of the different features to choose the most effective of them in a particular case [50]. The image processing algorithms that need to be implemented to extract those features will be discussed next.

4.3. Classification

Many classification techniques can be examined associated with the extracted features to evaluate the performance of each of them. The classification techniques aforementioned in the background section were chosen to be examined to identify the most optimal solution for this application. The extracted features were used to build and train the proposed classification models.

4.4. Cross Validation

The cross-validation technique was used for evaluating the performance validation of a specific model for a sample of data for future data. It was done by dividing the data into groups: the training set, which was used to apply the model, and the testing set, which was used to calculate the results error ratio. This will help improve the proposed model to obtain the highest performance.

5. Experiments

MATLAB software was used, including the Image processing toolbox, Graphic toolbox, Mathematic toolbox, and Statistics and ML toolbox to implement the proposed method, including all steps from preprocessing to evaluation and improving the performance.

In this section, we describe all the experiments that were done. It begins with describing the collected dataset and how it was preprocessed to test the proposed model. Then, we explain how we implemented the experiments and improved them until we obtained the best performance.

5.1. Dataset Description

We received ethical approval from the Standing Committee for Research Ethics on Living Creatures (SCRELC) at Imam Abdulrahman Bin Faisal University (IAU) (reference number IRB-PGS-2018-09-292). Then, the mammography dataset used in this study was collected from King Fahd Hospital of the University (KFUH) affiliated to IAU in Khobar,

Kingdom of Saudi Arabia. The collected data are new and have not been published before. We collected it from consecutive recent patients who had digital mammograms, considering the variety between the chosen samples in terms of density, shape, and tumor size. The collected dataset contains 42 instances, 21 benign (BIRADS 2) and 21 malignant (BIRADS 5). Each instance has 4 two dimensional images for the right and left breast. So, there were two different views of the images for each side (mediolateral–oblique (MLO) and cranial–caudal (CC)).

Whether the view of the image was MLO or CC, we selected the image that had a clear tumor in every instance to prepare the data for the tumor annotating step.

5.2. Data Annotation

By collaborating with the radiologist, the Segmenter tool in MATLAB was used in this step to manually annotate the tumor region by drawing its border in all the selected images, as seen in Figure 3.

After drawing the tumor border, the segmented tumor area (ROI) was converted to binary and masked images, as seen in Figure 4. The image conversion was done using the function generator by the Segmenter tool. The masked image represents only the ROI and it was extracted from the breast without its background. The binary and masked images were exported to use in the features extraction step.

5.3. Features Extraction

Table 1 identifies each extracted feature with the factor that describes it, and Figure 5 shows a sample of the extracted values from a few instances. The extracted features were selected based on the tumor density, shape, and texture.

Table 1. Extracted features.

Based	Feature	Description	Equation	Equation Number
Density	Mean	The mean is one of the most basic notions in experimental sciences. It is used in many applications of everyday life [51]. It is an average of the pixels' density value. Equation (1) is used to calculate the mean of ROI density and the mean of the whole image of breast density. Equation (2) is the calculated proportion of dissimilarity between the ROI mean and the image mean.	$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ X vector contains the value of n pixels where n represents the number of image pixels and X_i each pixel density.	(1)
			$Mean\ Contrast = \frac{ROI\ \bar{X}}{image\ \bar{X}}$	(2)
	SD	Standard deviation is used to measure the degree of dispersion and variation of the image [52]. This value represents the mean distance between the pixel and the mean. Equation (3) is the calculated SD (ROI) and SD (image). Equation (4) is the calculated Max (SD ROI).	$SD = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$	(3)
$Max\ SD\ (ROI) = Mean(ROI) + SD\ (ROI)$			(4)	
Histogram	It represents the distribution of density values of ROI by graphical bars. The sum of these values was calculated using Equation (5). Normalized of histogram values by resetting the density value were calculated by Equation (7).	$sum\ his\ ROI = \sum_{i=1}^n H_i$ where n refers to bars number and H_i refers to each bar value	(5)	
		$area\ ROI = N \times P^2$ N represents the number of pixels in ROI, and P is the pixel size.	(6)	
		$NHist = \frac{sum\ his\ ROI}{area\ ROI}$	(7)	

Table 1. Cont.

Based	Feature	Description	Equation	Equation Number
Shape	ROI Compactness	One of the shape circularity measurement techniques in digital images associated with the shape perimeter and its area [53–55]. It is calculated by Equation (8) [55].	$Compactness = \frac{Perimeter^2}{4 \times 3.14 \times N}$ <i>N</i> represents the number of pixels.	(8)
	Kurtosis	Kurtosis measures the tailed-level compared to a normal distribution that indicates whether the data looks flat or not [56,57]. It is calculated by Equation (9).	$Kurtosis = \frac{E(x-\bar{x})^4}{s^4}$ \bar{x} is the mean of <i>x</i> and <i>s</i> is SD of ROI.	(9)
Texture	Solidity	It illustrates the ratio of the pixels on the convex body in the ROI area by Equation (10) [58].	$Solidity = \frac{Area}{ConvexArea}$	(10)
	Correlation	It measures the correlative value between the pairs of pixels [48,59]. If the gray level between the pixel pair is stable in ROI, it indicates a high correlation value [48]. Equation (11) is used to calculate this feature.	$Correlation = \frac{\sum_{i=1}^n \sum_{j=1}^n (i-\bar{x})(j-\bar{x})p[i,j]}{\sigma^2}$ where \bar{x} is the mean and σ^2 is the SD	(11)
	Variance	It is a squared of SD. It is used to measure the degree of dispersion and the variation in the pixels value in the ROI [52,60].	$Variance = S^2$	(12)
	Uniformity	In this feature, we make the density in the ROI's histogram uniform by squaring the normalized histogram value.	$Uniformity = NHist^2$	(13)

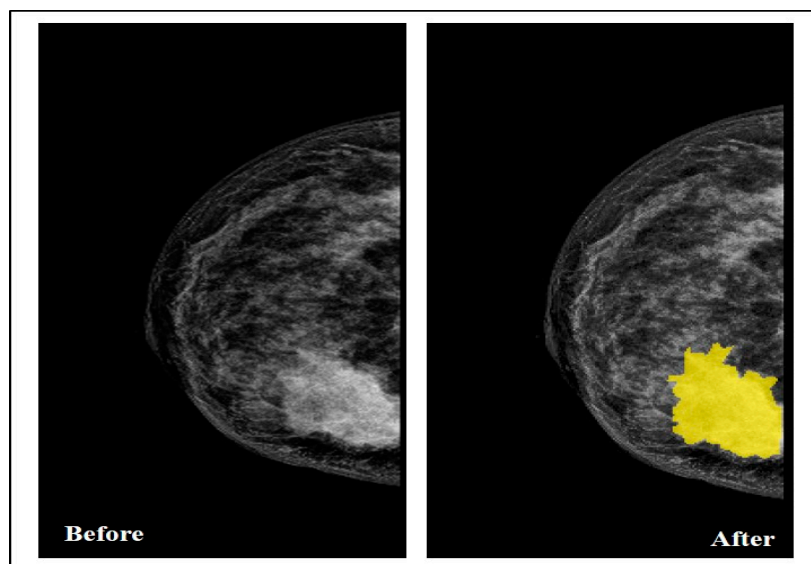


Figure 3. Breast before and after drawing the tumor border.

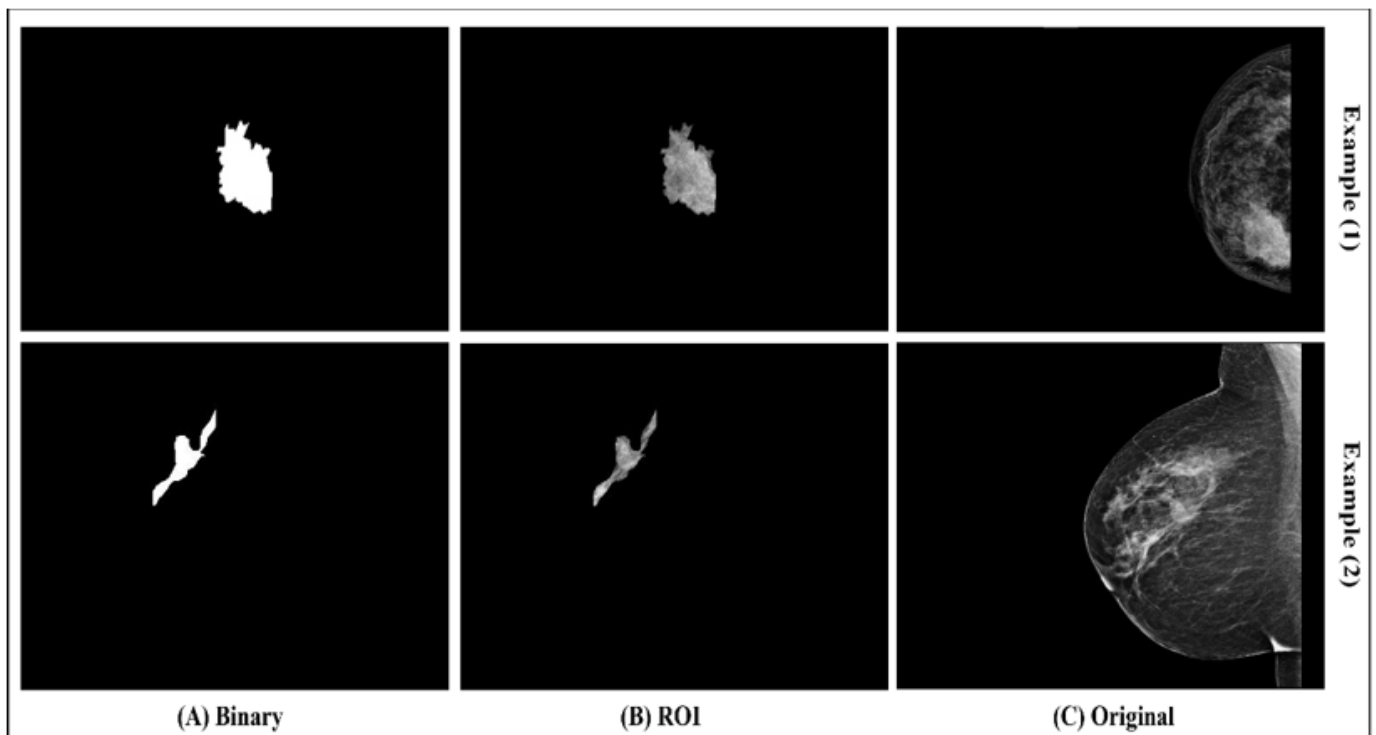


Figure 4. (A) Binary image; (B) ROI; (C) original image.

1	2	3	4	5	6	7	8	9	10	11	12	13
meanROI	Contrast	stdROI	stdimage	maxstd	compactnessROI1	solidity	kurtosiss	variance	sumhistROI	uniformity	correlationROI	label
123.2667	0.6118	35.7594	42.9431	159.0261	0.8769	0.9375	2.1291	1.2787e+03	6	740.4322	7.1310	'Benign.jpg'
111.2453	0.6487	26.2476	35.0165	137.4929	1.9239	0.7681	2.7665	688.9376	15	370.6774	9.7152	'Malignant.j...
198.7813	0.3510	78.0930	37.6603	276.8742	1.3085	0.8205	2.8541	6.0985e+03	6	1.4642e+03	3.2653	'Benign.jpg'
147.4826	0.4162	43.0050	33.0625	190.4876	0	1	2.4662	1.8494e+03	45	8.4340e+07	5.9295	'Malignant.j...
121.4783	0.5681	29.2173	35.1509	150.6955	1.1492	0.8313	2.3070	853.6488	12	1.2597e+03	8.7277	'Benign.jpg'
137.5393	0.4833	30.0066	37.7555	167.5459	2.0782	0.7982	2.5659	900.3953	21	579.7042	8.4981	'Malignant.j...
117.9720	0.6035	27.3517	43.9766	145.3237	1.3288	0.8412	3.4625	748.1159	15	114.5667	9.3230	'Benign.jpg'
166.0088	0.5304	39.3318	50.5761	205.3406	4.3445	0.5571	2.8850	1.5470e+03	15	29.2521	6.4825	'Malignant.j...

Figure 5. Samples of the extracted features from a few cases.

5.4. Experiments and Results

Several experiments on the classifiers mentioned above have improved the models' performance based on the input dataset. The experiments started with the default models' properties tuned by functions developed in MATLAB with all the extracted features. These experiments continued until we optimized all the models' properties and achieved the highest accuracy by integrating one of the features selection functions, which will be explain in detail in the following sub-sections.

5.4.1. First Experiment

The mentioned classification techniques with default properties mode were tested with all features for the model's training. Two cross-validation (CV) methods were used for the dataset partitioning: Holdout and K-Fold.

This experiment obtained different results for each classifier, as seen in Figure 6, with the two different CV partitioning methods creating different training sets. These variations are due to different factors depending on the used classifier. However, the used classifiers with default properties and random selection of the training sets using either Holdout or

K-Fold are inefficient. We could not trust the results even if we obtained high performance, which needs improvements.

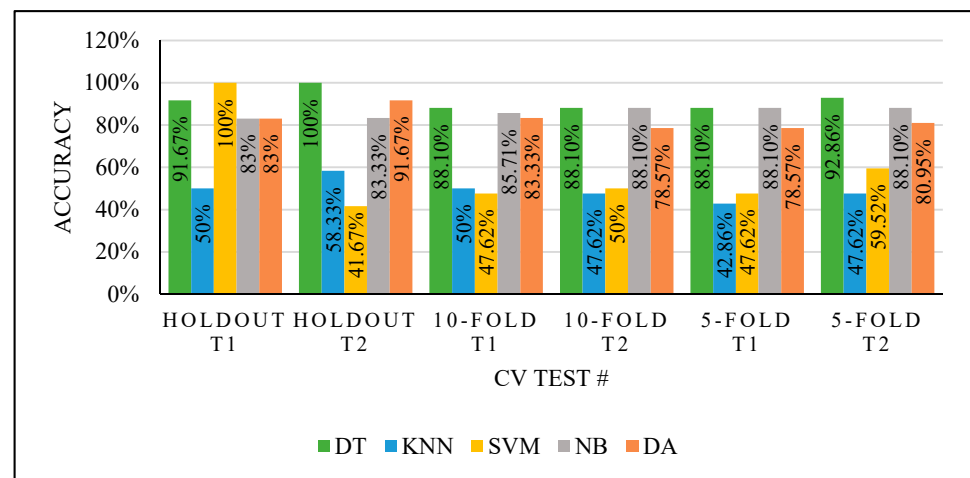


Figure 6. Experiment 1 results.

5.4.2. Second Experiment

Various range criteria were used for manual dataset partitioning with all the extracted features to train and test the classifiers, where each set has a balanced number of benign and malignant cases to improve the models' performance. The dataset partitioning range categorized the following tests. Only KNN parameters were modified in this experiment because it produced poor performance in all the Experiment 1 tests. The KNN parameters were modified after testing all the distances with 1 to 10 K values with two different partitioning ranges, 70–30% and 75–25%.

Based on these tests, the best parameters' values were tuned, as shown in Table 2. Euclidean distance performed very well with different values of K. The selected K value was decided by considering that the small values are more flexible for a fit, but it will increase the variation. The large values use high computational power; thus, the selected value was not very small but also not the largest tested value. The number of neighbors in Table 2 was selected to avoid those possibilities.

Table 2. Modified KNN parameters.

Properties	Value
Number of neighbors	5
Distance	Euclidean

In this experiment, we tried to improve all testing classifiers by guaranteeing that the sets after the partitioning process contains balanced data. Four partitioning ranges were used; each range was tested five times, wherein each test, we used a different part from the dataset as a testing set. Figure 7 shows how the split process for the testing set was done.

All the test results produced the best performance when using the end of the dataset (C) as the testing set, as seen in Figures 8–11. This finding is due to the variety of cases in the training set. We did not manage or control the dataset arrangement, so if the cases used to train the model are similar, it will fail to diagnose other different cases. It means the cases from the beginning of the dataset vary, making the classification process more efficient and accurate. So, the used model was trained with enough variation in the cases to classify the new cases accurately.

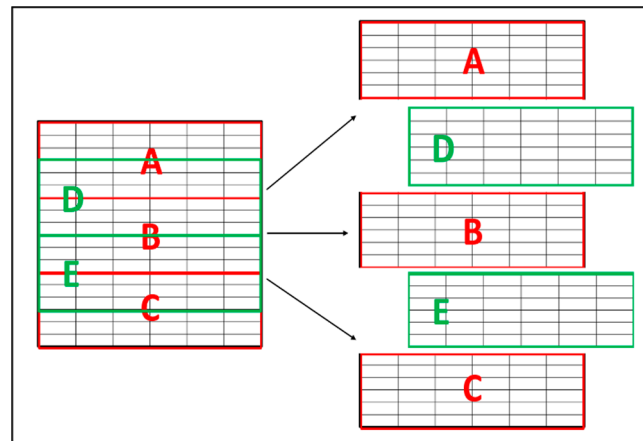


Figure 7. Testing set selection.

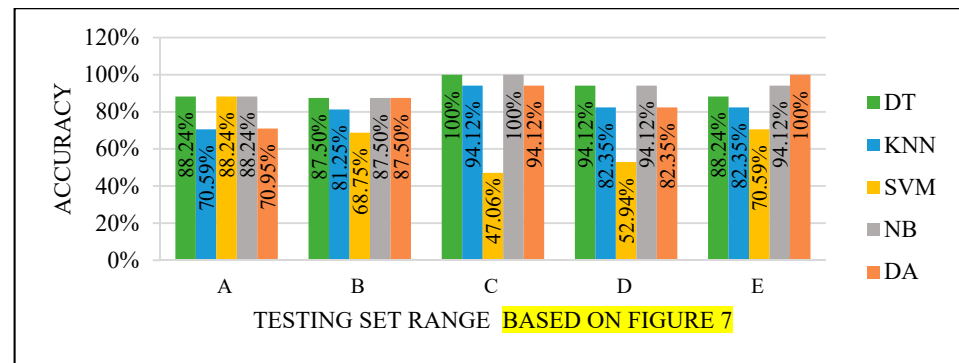


Figure 8. Models' performance with the dataset splitting criteria at 60% training and 40% testing.

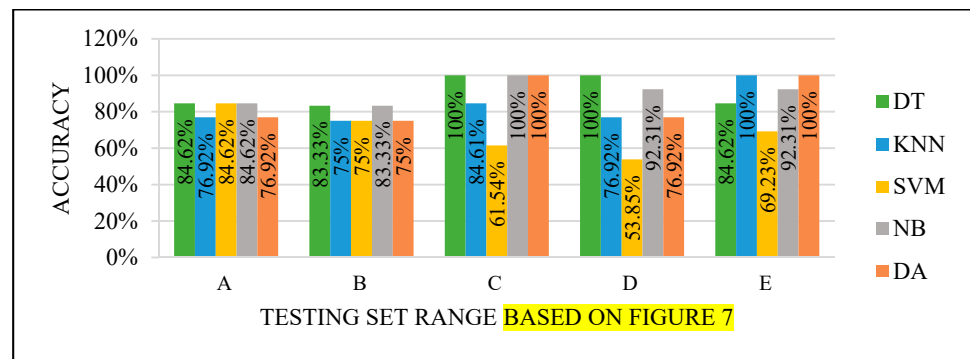


Figure 9. Models' performance with the dataset splitting criteria at 70% training and 30% testing.

SVM did not perform well, and this could be due to two factors: model properties or data dimensionality. So, it needed to be improved to achieve the highest accuracy efficiently.

All the models needed to be improved to guarantee the results that the end-user can trust with the least computation time and least resource consumption. For this purpose, we integrated two optimization functions in the third experiment that could achieve the best performance, which are explained in the following.

Forward Sequential Feature Selection (FSFS)

Its name refers to the repeated nature of the process [61] until obtaining the best feature to achieve accurate classification results. It has a faster computational time where it reduces the needed calculation process of a large number of attributes and saves the computing resources such as memory [61,62].

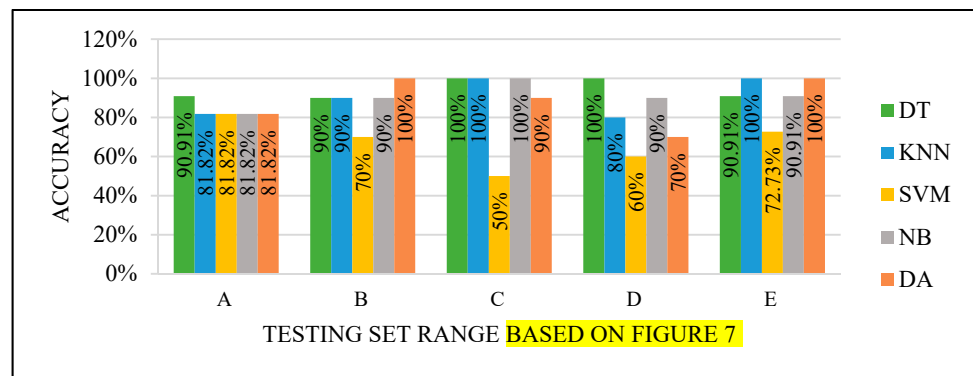


Figure 10. Models' performance with the dataset splitting criteria at 75% training and 25% testing.

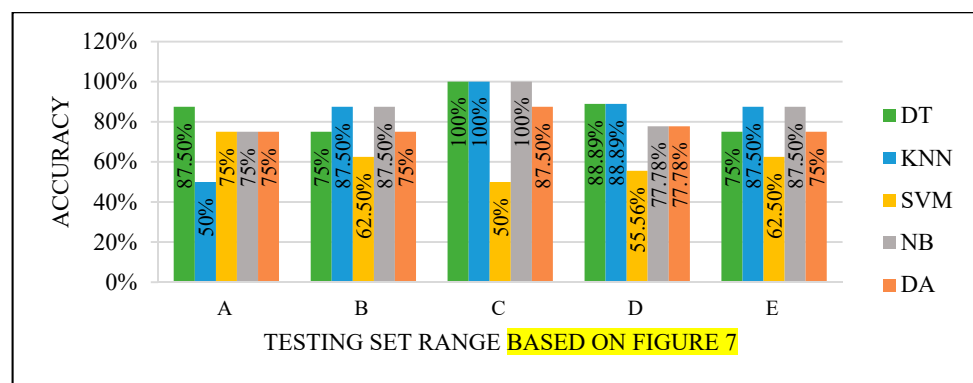


Figure 11. Models' performance with the dataset splitting criteria at 80% training and 20% testing.

Hyperparameters Optimization Function

Most of the classification techniques' performance depends on the settings of the hyper-parameters of the classifiers [63,64].

In recent studies, Bayesian optimization proved that it performs better than other methods [63,64]. It works based on an iterative Gaussian process, a robust technique for ML [65].

5.4.3. Third Experiment

We combined two functions to improve the classifiers' accuracy and computational time to obtain an efficient result. The FSFS and hyper-parameters optimization functions were used together.

Cases were selected for the training and testing partitions randomly and manually, resulting in 70% of the dataset for training and 30% for testing. The random process was repeated three times to test the performance with as many possible differences of the selected cases.

Random Selection

The dataset was reordered randomly using a function in MATLAB. Then it was partitioned to 70% of the dataset for training and the rest for testing. We repeated this process three times. Every time the test created training and testing sets that contained different cases due to reordering them using the random function. Thus, the aim of the many tests was to experiment with the performance with different sets that contain randomly selected cases. For each created training and testing set, the models were trained using a 10-fold CV, and then tested using a hyper-parameter function to optimize the parameters with all classifiers.

The steps mentioned above were used with the FSFS function to select the best feature (BF) and the best two features (B2F) again.

We achieved high performance by using randomly selected cases to split the dataset for the training and testing phase and using parameter optimization and feature selection functions that reduced the data dimensionality, as seen in Figure 12. This means the models that needed to be trained with balanced data to perform well, such as DT, can perform very well with unbalanced data once using parameters optimization and dimensionality reduction. However, while we could not control the selected cases variation for training, the results were not efficient enough, and we could not guarantee the performance of the classifiers with new cases.

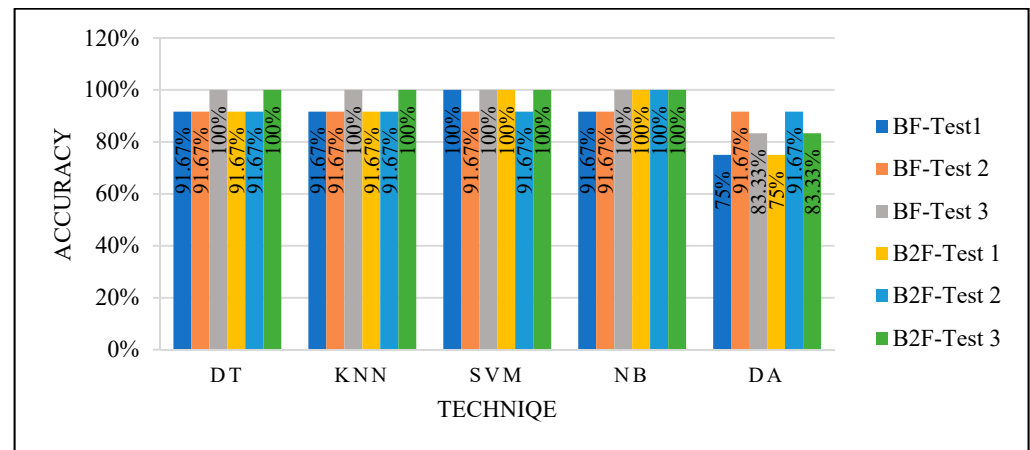


Figure 12. Models' performance with a random selection of the training set.

Manual Selection

The dataset manually selected 70% of cases for training and 30% for testing to guarantee a balance and variation in the training cases. The training and testing sets were used to test all the classifiers with the FSFS and parameter optimization functions.

They were tested two times: by selecting the best feature and the best two features. Figure 13 shows the classifiers' performance with both the best and best two features. All the classifiers produced an excellent performance.

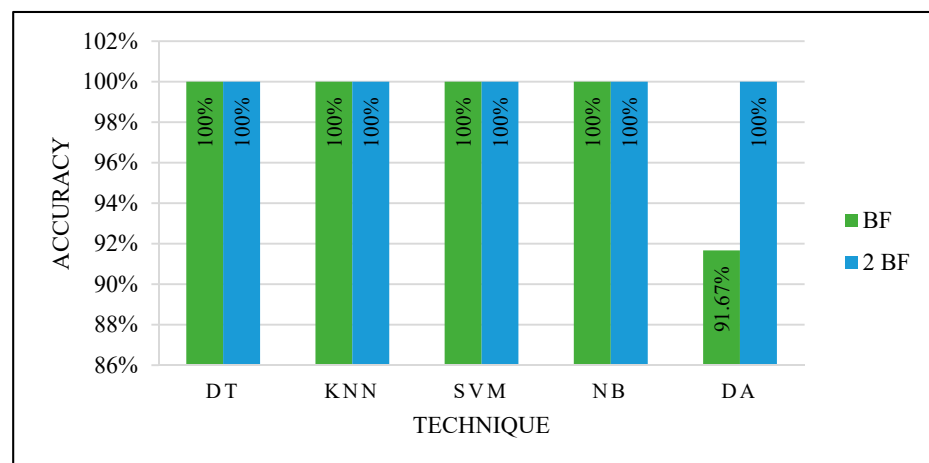


Figure 13. Models' performance with manual selection of the training set.

We achieved the highest accuracy using manually selected cases to split the dataset for the training and testing phase, the parameter optimization function, and the feature selection function that reduced the data dimensions. This test considered all the factors that may affect the classifiers' performance. The selected cases were balanced and varied. All the classifiers were modified to produce the optimized parameters, and the dimensions were reduced.

Based on all the previous experiments where the manual selection of the cases for the training set achieved excellent performance with the best feature and the best two features selection, the proposed method in the previous section shown in Figure 2 was optimized, as shown in Figure 14.

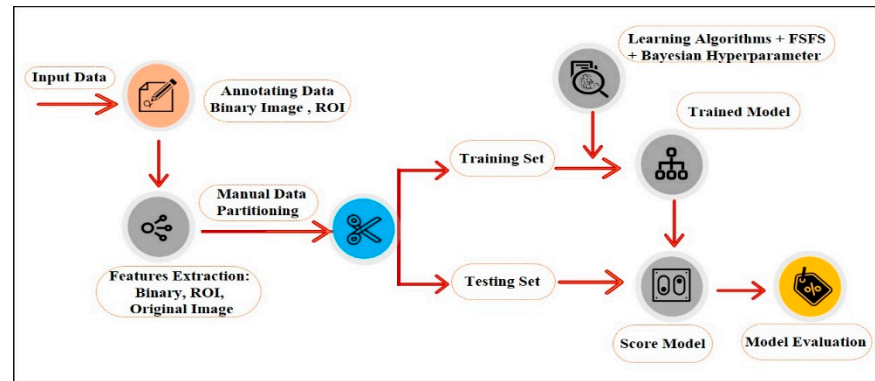


Figure 14. Optimized method.

We considered the computational time an essential factor for the end-user when they want to predict a new case using one of the optimized techniques. Predicting one or two extracted features from one image instead of 12 extracted features from three images (original image, ROI, and binary) of one instance will save the needed computation time and computer resources. So, the least features were tested. Once the test performed very well, it means the training and testing sets with the best feature for all the classifiers, except DA, which has an excellent performance with two features, are the most efficient, and meet our expectations. Based on the results, the most crucial feature was the ROI compactness. In general, the shape features gave an excellent performance compared to the other features.

5.4.4. Proving Test

To prove the optimized models' performance with the best features, 13 unlabeled mammograms were collected from KFUH. The collected cases were processed to extract the features and tested with each optimized classifier as in the third experiment with manual partitioning. After the prediction with the optimized models, the results were shown to the radiologist at KFUH to check the correctness of the optimized models. Figure 15 shows the results.

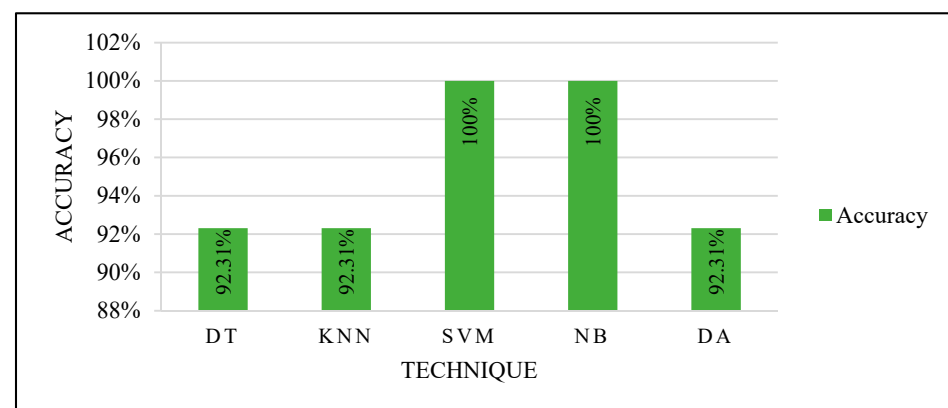


Figure 15. Proving the test results.

Based on the proving test results, SVM and NB are the most powerful, efficient, and accurate ML techniques. SVM and NB proved their efficiency to solve this problem. Although other techniques performed very well, SVM and NB are the most accurate techniques. In addition, they were able to accurately predict using only one feature: the

ROI compactness. This indicates the ability to deal with only one feature that could save the computation time to calculate several features for every input case.

6. Discussion

As explained in the previous section, after extracting 12 features from the annotated mammography dataset, several experiments were done with various criteria of partitioning and hyper-parameter optimizations to obtain the highest performance with all the tested classifiers. We compared to determine the best techniques to trust its results. Four factors may affect the performance of the classifiers: data balance, training data variety, data dimensionality, and the used classifiers' parameters. These factors depend on each classifier.

In the first experiment, we used two CV techniques and tested them several times with all classifiers. The results produced while using the classifiers for each test in this experiment or when using two tests for each classifier indicated that the classifiers' performances were not stable. That led us to do more experiments with a managed partitioning process to improve the performance of the classifiers.

The second experiment was done by managing the data partitioning process based on the first experiment's results. Four ranges of dataset splitting were assessed; each range was tested five times where each time we picked the testing set from a different place of the dataset. This experiment showed good performance for DT, NB, and DA. SVM, which is one of the most powerful techniques, was not performing well. So, we planned more experiments to improve its performance and improve the computation time and resources consumption of all the classifiers.

We combined Forward Sequential Feature Selection and Bayesian hyper-parameter optimization functions to improve the results and guarantee the performance efficiency of the tested classifiers. These functions were used with random and manual training and testing sets partitioning. Both were tested by selecting the best feature and the best two features. The test with only one and two features was selected due to needing the least computation time and least resources consumption. This combination gave an excellent performance with 100% accuracy with manual partition because of the various cases selection for the training set, which is an essential factor.

Figures 16–19 show how the performance of each technique was changed among all the experiments. The parameters of NB were not changed, and it worked very well once it was combined with the feature selection function.

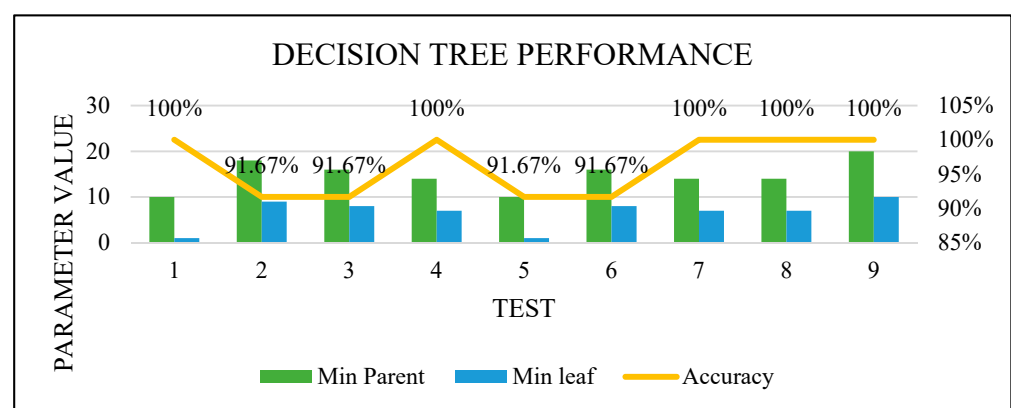


Figure 16. DT optimization.

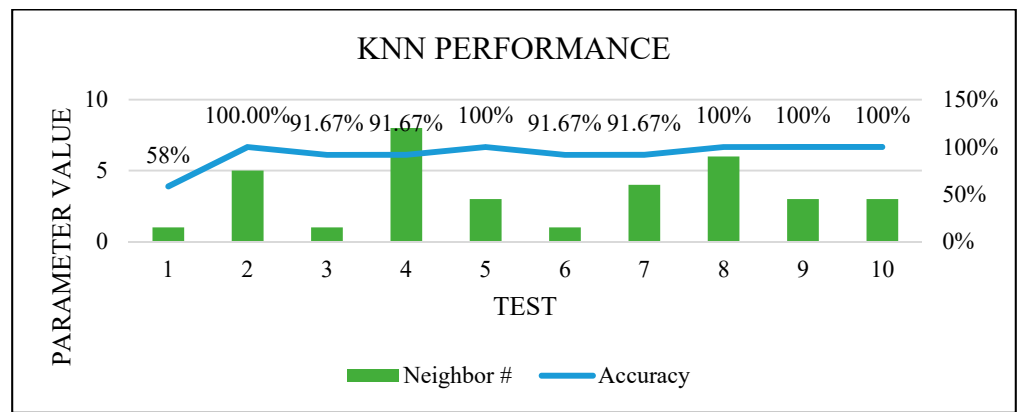


Figure 17. KNN optimization.

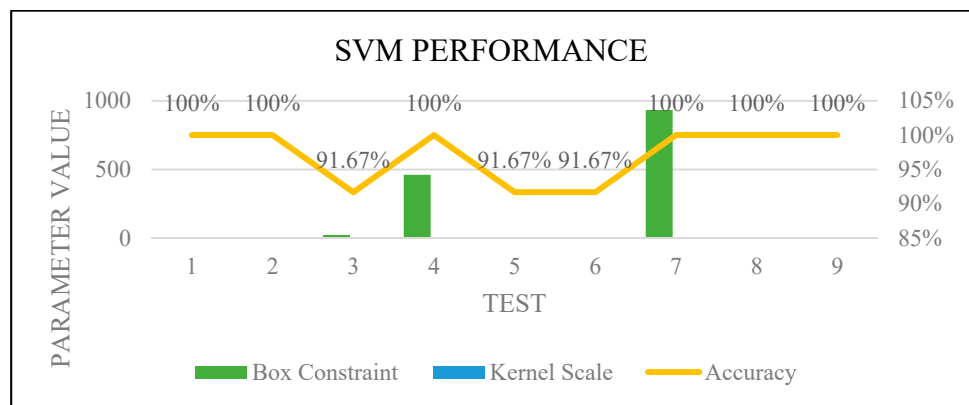


Figure 18. SVM optimization.

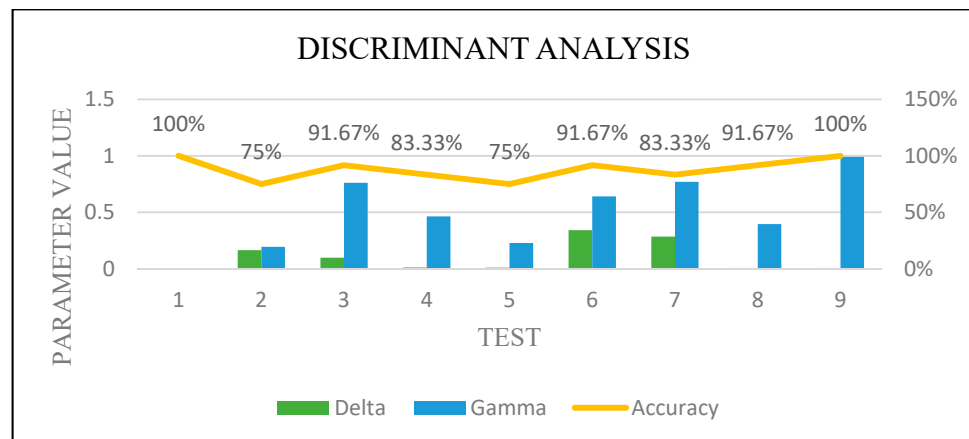


Figure 19. DA optimization.

The newly trained models with the best features, manual partition, and parameters optimization were used in proving the test using unlabeled data, to be confirmed by the radiologist that the optimized models obtained the aimed performance.

Based on all experiments findings, the proposed ML-based CAD system is shown in Figure 20. It achieved the highest accuracy with the least computation time and the least resources consumption for input processing. According to the proven test results shown in Figure 15, the optimized models used in the third experiment by selecting the best feature and manually partitioning the dataset led to using the optimized parameters for SVM and NB.

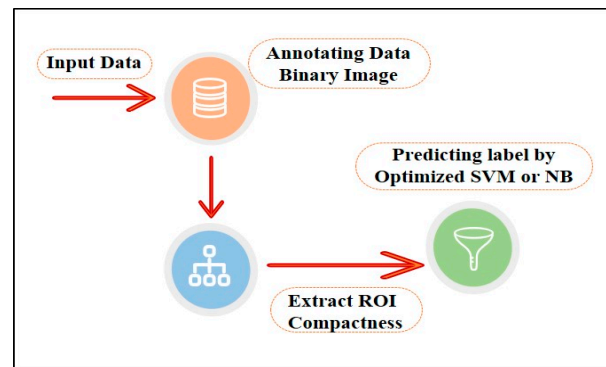


Figure 20. Recommended model.

The proposed method for a CAD system can be performed using the following steps. The mammogram needs to be annotated and saved as a binary image of the ROI manually. Then, one should calculate the ROI compactness using Equation (8) in Table 1, where it should be enough, as we proved in the third experiment. Then, based on the proven test results, we recommend performing the classification process by either SVM or NB with its optimized parameters, which resulted in 100% accuracy for the classification process of new cases.

7. Conclusions and Future Work

In this study, we developed a breast cancer diagnosis based on an ML method. This method is considered to avoid the possible errors in tumor segmentation by current CAD systems. Errors may occur due to the complex nature of breast tissue. The proposed method uses the radiologist's knowledge to draw the tumor border using a touchpad. The radiologist's rule in the proposed method makes the process more accurate and efficient by considering the border smoothness, tailedness, or coarseness. After annotating the ROI, we saved the manually annotated area as a binary image (black and white) and masked image, which presented the tumor only (without its background by showing its margins in black). This was used to extract the ROI features related to density, shape, and texture. Then, we built the diagnostic model to work while considering the least computation time and saving the computer resources.

This study concludes that SVM and NB with tuned parameters are the most powerful and efficient classification techniques based on the ROI compactness, which we extracted using the binary image of ROI, thus achieving the maximum efficiency and resource conservation. We recommend doing more tests with a large number of samples for training the recommended model. In addition, further testing for the other classifiers, either hybrid or individual, is recommended.

Author Contributions: Conceptualization, M.M.A. and J.A.; data curation, M.M.A.; methodology, M.M.A.; project administration, J.A.; resources, A.A.; validation, M.M.A. and A.A.; supervision, J.A.; writing—original draft, M.M.A.; writing—review and editing, J.A. All authors will be informed about each step of manuscript processing, including submission, revision, revision reminder, etc., via emails from our system or the assigned Assistant Editor. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was conducted according to the guidelines of the Declaration of Standing Committee for Research Ethics on Living Creatures (SCRELC) at Imam Abdulrahman Bin Faisal University (IAU), and approved by the Institutional Review Board at IAU by reference number IRB-PGS-2018-09-292, On 19 December 2018.

Informed Consent Statement: In this study, patient consent was waived since the data used does not contain any patient data, and this study hadn't contact directly with any patient in any way, and

the used data was collected from King Fahd Hospital of the University (KFUH) database, affiliated to IAU in Khobar, Kingdom of Saudi Arabia.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available since it was obtained after the approval from Institutional Review Board as mentioned above.

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

References

1. World Health Organization. Cancer. Available online: <https://www.who.int/news-room/fact-sheets/detail/cancer> (accessed on 21 November 2021).
2. World Health Organization. *WHO | Breast Cancer*; WHO: Geneva, Switzerland, 2021.
3. Health Days 2018—World Cancer Day. 2018. Available online: <https://www.moh.gov.sa/en/HealthAwareness/HealthDay/2018/Pages/HealthDay-2018-02-04.aspx> (accessed on 12 April 2019).
4. Health Days 2018—Breast Cancer Awareness Month. 2018. Available online: <https://www.moh.gov.sa/en/HealthAwareness/HealthDay/2018/Pages/HealthDay-2018-10-01-31.aspx> (accessed on 12 April 2019).
5. Dromain, C.; Boyer, B.; Ferré, R.; Canale, S.; Delalogue, S.; Balleyguier, C. Computed-aided diagnosis (CAD) in the detection of breast cancer. *Eur. J. Radiol.* **2013**, *82*, 417–423. [[CrossRef](#)]
6. Witten, I.H.; Frank, E.; Hall, M.A. Algorithms: The Basic Methods. In *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed.; Morgan Kaufmann: Burlington, MA, USA, 2011; pp. 85–145.
7. Tsiligaridis, J. A Criteria based Decision Tree for Classification. In Proceedings of the Central European Conference on Information and Intelligent Systems, Varaždin, Croatia, 18–20 September 2013; pp. 14–18.
8. Kotsiantis, S.B. Decision trees: A recent overview. *Artif. Intell. Rev.* **2013**, *39*, 261–283. [[CrossRef](#)]
9. Cherif, W. Optimization of K-NN algorithm by clustering and reliability coefficients: Application to breast-cancer diagnosis. *Procedia Comput. Sci.* **2018**, *127*, 293–299. [[CrossRef](#)]
10. Diz, J.; Marreiros, G.; Freitas, A. Applying Data Mining Techniques to Improve Breast Cancer Diagnosis. *J. Med. Syst.* **2016**, *40*, 203. [[CrossRef](#)] [[PubMed](#)]
11. Castella, C.; Kinkel, K.; Eckstein, M.P.; Sottas, P.E.; Verdun, F.R.; Bochud, F.O. Semiautomatic Mammographic Parenchymal Patterns Classification Using Multiple Statistical Features. *Acad. Radiol.* **2007**, *14*, 1486–1499. [[CrossRef](#)]
12. Thulin, J. Machine Learning-Based Classifiers in the Direkt Profil Grammatical Profiling System. Master’s Thesis, Lund University, Lund, Sweden, 2007.
13. Olson, D.; Delen, D. Rough Sets. In *Advanced Data Mining Techniques*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 87–109.
14. Cui, Y. Discriminant Analysis. In *Encyclopedia of Research Design*; Salkind, N.J., Ed.; SAGE Publications, Inc.: Thousand Oaks, CA, USA, 2010; pp. 371–375.
15. Nilashi, M.; Ibrahim, O.; Ahmadi, H.; Shahmoradi, L. A knowledge-based system for breast cancer classification using fuzzy logic method. *Telemat. Inform.* **2017**, *34*, 133–144. [[CrossRef](#)]
16. Kubat, M. *An Introduction to Machine Learning Communications Systems*; Springer: Berlin/Heidelberg, Germany, 2017.
17. Tang, J.; Rangayyan, R.M.; Xu, J.; El Naqa, I.; Yang, Y. Computer-Aided Detection and Diagnosis of Breast Cancer with Mammography: Recent Advances. *IEEE Trans. Inf. Technol. Biomed.* **2009**, *13*, 236–251. [[CrossRef](#)] [[PubMed](#)]
18. Jalalian, A.; Mashohor, S.B.T.; Mahmud, H.R.; Sariipan, M.I.B.; Ramli, A.R.B.; Karasfi, B. Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: A review. *Clin. Imaging* **2013**, *37*, 420–426. [[CrossRef](#)]
19. Kele, A.; Kele, A.; Yavuz, U. Expert system based on neuro-fuzzy rules for diagnosis breast cancer. *Expert Syst. Appl.* **2011**, *38*, 5719–5726. [[CrossRef](#)]
20. Costaridou, L. *Biomedical Image Processing*; Springer: Berlin/Heidelberg, Germany, 2011.
21. Labuda, N.; Seeliger, J.; Gedrande, T.; Kozak, K. Selecting Adaptive Number of Nearest Neighbors in k-Nearest Neighbor Classifier Apply Diabetes Data. *J. Math. Stat. Sci.* **2017**, *2017*, 1–13.
22. Olatunji, S.O. Improved email spam detection model based on support vector machines. *Neural Comput. Appl.* **2017**, *31*, 691–699. [[CrossRef](#)]
23. Akande, K.O.; Owolabi, T.O.; Olatunji, S.O. Investigating the effect of correlation-based feature selection on the performance of support vector machines in reservoir characterization. *J. Nat. Gas Sci. Eng.* **2015**, *22*, 515–522. [[CrossRef](#)]
24. Xin, C. Machine Learning Techniques for Cervigram Image Analysis. Master’s Thesis, Lehigh University, Bethlehem, PA, USA, 2016; p. 72.
25. Karamzadeh, S.; Abdullah, S.M.; Halimi, M.; Shayan, J.; Rajabi, M.J. Advantage and drawback of support vector machine functionality. In Proceedings of the 2014 International Conference on Computer, Communications, and Control Technology (I4CT), Langkawi, Malaysia, 2–4 September 2014; pp. 63–65.
26. Kitanovski, I.; Jankulovski, B.; Dimitrovski, I.; Loskovska, S. Comparison of feature extraction algorithms for mammography images. In Proceedings of the 2011 4th International Congress on Image and Signal Processing, Shanghai, China, 15–17 October 2011; Volume 2, pp. 888–892.

27. Kotsiantis, S.B. Supervised Machine Learning: A Review of Classification Techniques. In *Emerging Artificial Intelligence Applications in Computer Engineering: Real World AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*; IOS Press: Amsterdam, The Netherlands, 2007; pp. 3–24.
28. Han, J.; Kamber, M. *Data Mining: Concepts and Techniques*; Elsevier: Berlin/Heidelberg, Germany, 2006.
29. Yang, Z.R. Biological applications of support vector machines. *Brief. Bioinform.* **2004**, *5*, 328–338. [[CrossRef](#)]
30. Tay, F.E.-H.; Shen, L.; Cao, L. Application of Support Vector Machines in Financial Time Series Forecasting. *Ordinary Shares Exot. Methods* **2010**, *29*, 111–129.
31. Huang, S.; Cai, N.; Pacheco, P.P.; Narrandes, S.; Wang, Y.; Xu, W. Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. *Cancer Genom. Proteom.* **2018**, *15*, 41–51.
32. Raschka, S. Naive Bayes and Text Classification I—Introduction and Theory. *arXiv* **2014**, arXiv:1410.5329, 1–20.
33. Mohammadi, M.; Raahemi, B.; Akbari, A.; Nassersharif, B.; Moeinzadeh, H. Improving linear discriminant analysis with artificial immune system-based evolutionary algorithms. *Inf. Sci.* **2012**, *189*, 219–232. [[CrossRef](#)]
34. McLachlan, G.J. Discriminant analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2012**, *4*, 421–431. [[CrossRef](#)]
35. Şahan, S.; Polat, K.; Kodaz, H.; Güneş, S. A new hybrid method based on fuzzy-artificial immune system and k-nn algorithm for breast cancer diagnosis. *Comput. Biol. Med.* **2007**, *37*, 415–423. [[CrossRef](#)]
36. Akay, M.F. Support vector machines combined with feature selection for breast cancer diagnosis. *Expert Syst. Appl.* **2009**, *36*, 3240–3247. [[CrossRef](#)]
37. Ghosh, S.; Biswas, S.; Sarkar, D.; Sarkar, P.P. A novel Neuro-fuzzy classification technique for data mining. *Egypt. Inform. J.* **2014**, *15*, 129–147. [[CrossRef](#)]
38. Pak, F.; Kanan, H.R.; Alikhassi, A. Breast cancer detection and classification in digital mammography based on Non-Subsampled Contourlet Transform (NSCT) and Super Resolution. *Comput. Methods Programs Biomed.* **2015**, *122*, 89–107. [[CrossRef](#)]
39. Xie, W.; Li, Y.; Ma, Y. Breast mass classification in digital mammography based on extreme learning machine. *Neurocomputing* **2016**, *173*, 930–941. [[CrossRef](#)]
40. Aličković, E.; Subasi, A. Breast cancer diagnosis using GA feature selection and Rotation Forest. *Neural Comput. Appl.* **2017**, *28*, 753–763. [[CrossRef](#)]
41. Wang, H.; Zheng, B.; Yoon, S.W.; Ko, H.S. A support vector machine-based ensemble algorithm for breast cancer diagnosis. *Eur. J. Oper. Res.* **2018**, *267*, 687–699. [[CrossRef](#)]
42. Joshi, A.; Mehta, A. Analysis of K-Nearest Neighbor Technique for Breast Cancer Disease Classification. *Int. J. Recent Sci. Res.* **2018**, *9*, 26126–26130.
43. Liu, N.; Qi, E.S.; Xu, M.; Gao, B.; Liu, G.Q. A novel intelligent classification model for breast cancer diagnosis. *Inf. Process. Manag.* **2019**, *56*, 609–623. [[CrossRef](#)]
44. Chaurasia, V.; Pal, S. Applications of Machine Learning Techniques to Predict Diagnostic Breast Cancer. *SN Comput. Sci.* **2020**, *1*, 270. [[CrossRef](#)]
45. Abdar, M.; Zomorodi-Moghadam, M.; Zhou, X.; Gururajan, R.; Tao, X.; Barua, P.D.; Gururajan, R. A new nested ensemble technique for automated diagnosis of breast cancer. *Pattern Recognit. Lett.* **2020**, *132*, 123–131. [[CrossRef](#)]
46. Jiang, Y. Classification of Breast Lesions from Mammograms. In *Handbook of Medical Image Processing and Analysis*; Academic Press: Cambridge, MA, USA, 2008.
47. Haralick, R.M.; Shanmugam, K.; Dinstein, I. Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *6*, 610–621. [[CrossRef](#)]
48. Mahmood, F.H.; Abbas, W.A. Texture Features Analysis using Gray Level Co-occurrence Matrix for Abnormality Detection in Chest CT Images. *Iraqi J. Sci.* **2016**, *57*, 279–288.
49. Tahir, M.A.; Bouridane, A.; Kurugollu, F. An FPGA Based Coprocessor for GLCM and Haralick Texture Features and their application in prostate cancer classification. *Analog Integr. Circuits Signal Process.* **2005**, *43*, 205–215. [[CrossRef](#)]
50. Turtinen, M. Learning and Recognizing Texture Characteristics Using Local Binary Patterns. Ph.D. Thesis, University of Oulu, Oulu, Finland, 2007.
51. Sarkar, J.; Rashid, M. Visualizing Mean, Median, Mean Deviation, and Standard Deviation of a Set of Numbers. *Am. Stat.* **2016**, *70*, 304–312. [[CrossRef](#)]
52. Bland, J.M.; Altman, D.G. Statistics notes: Measurement error. *BMJ* **1996**, *312*, 1654. [[CrossRef](#)] [[PubMed](#)]
53. Bribiesca, E. An easy measure of compactness for 2D and 3D shapes. *Pattern Recognit.* **2008**, *41*, 543–554. [[CrossRef](#)]
54. Li, W.; Goodchild, M.F.; Church, R.L. An Efficient Measure of Compactness for 2D Shapes and its Application in Regionalization Problems. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 1227–1250. [[CrossRef](#)]
55. Montero, R.S.; Bribiesca, E. State of the Art of Compactness and Circularity Measures. *Int. Math. Forum* **2009**, *4*, 1305–1335.
56. Kim, H.-Y. Statistical notes for clinical researchers: Assessing normal distribution (2) using skewness and kurtosis. *Restor. Dent. Endod.* **2013**, *38*, 52. [[CrossRef](#)]
57. Dorić, D.; Nikolić-Dorić, E.; Jevremović, V.; Mališić, J. On measuring skewness and kurtosis. *Qual. Quant.* **2009**, *43*, 481–493. [[CrossRef](#)]
58. Measure Properties of Image Regions—MATLAB Regionprops. MathWorks China. Available online: <https://ww2.mathworks.cn/help/images/ref/regionprops.html#description> (accessed on 22 March 2019).

59. Zayed, N.; Elnemr, H.A. Statistical Analysis of Haralick Texture Features to Discriminate Lung Abnormalities. *Int. J. Biomed. Imaging* **2015**, *2015*, 12. [[CrossRef](#)]
60. Hozo, S.P.; Djulbegovic, B.; Hozo, I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med. Res. Methodol.* **2005**, *5*, 13. [[CrossRef](#)]
61. Chandrashekar, G.; Sahin, F. A survey on feature selection methods. *Comput. Electr. Eng.* **2014**, *40*, 16–28. [[CrossRef](#)]
62. Marcano-Cedeño, A.; Quintanilla-Domínguez, J.; Cortina-Januchs, M.G.; Andina, D. Feature selection using Sequential Forward Selection and classification applying Artificial Metaplasticity Neural Network. In Proceedings of the IECON 2010—36th Annual Conference on IEEE Industrial Electronics Society, Glendale, AZ, USA, 7–10 November 2010; pp. 2845–2850.
63. Eggenesperger, K.; Feurer, M.; Hutter, F. Towards an empirical foundation for assessing Bayesian optimization of hyperparameters. In Proceedings of the NIPS Workshop on Bayesian Optimization in Theory and Practice, Lake Tahoe, NV, USA, 10 December 2013; pp. 1–5.
64. Snoek, J.; Larochelle, H.; Adams, R.P. Practical Bayesian Optimization of Machine Learning Algorithms. *arXiv* **2012**, arXiv:1206.2944, 1–9.
65. Tamura, R.; Hukushima, K. Bayesian optimization for computationally extensive probability distributions. *PLoS ONE* **2018**, *13*, e0193785. [[CrossRef](#)]