Review Article

Tabu Search and Machine-Learning Classification of Benign and Malignant Proliferative Breast Lesions

Habib Dhahri [ID],[1,2] **Ines Rahmany** [ID],[2] **Awais Mahmood** [ID],[1] **Eslam Al Maghayreh** [ID],[3]
and Wail Elkilani[1]

[1]*College of Applied Computer Sciences (CACS), Al-Muzahimiyah Branch, King Saud University, Riyadh, Saudi Arabia*
[2]*Faculty of Sciences and Technology of Sidi Bouzid, University of Kairouan, Kairouan, Tunisia*
[3]*Computer Science Department, Yarmouk University, Irbid, Jordan*

Correspondence should be addressed to Habib Dhahri; hdhahri@ksu.edu.sa

Breast cancer is the most diagnosed cancer among women around the world. The development of computer-aided diagnosis tools is essential to help pathologists to accurately interpret and discriminate between malignant and benign tumors. This paper proposes the development of an automated proliferative breast lesion diagnosis based on machine-learning algorithms. We used Tabu search to select the most significant features. The evaluation of the feature is based on the dependency degree of each attribute in the rough set. The categorization of reduced features was built using five machine-learning algorithms. The proposed models were applied to the BIDMC-MGH and Wisconsin Diagnostic Breast Cancer datasets. The performance measures of the used models were evaluated owing to five criteria. The top performing models were AdaBoost and logistic regression. Comparisons with others works prove the efficiency of the proposed method for superior diagnosis of breast cancer against the reviewed classification techniques.

1. Introduction

Breast cancer is a common cause of death and is the type of cancer that is widespread among women worldwide [1]. Many imaging techniques and tools have been developed for early detection and treatment of breast cancer and for reducing the number of deaths as a result of it [2]. There have been many breast cancer diagnosis methods that were used to increase diagnostic accuracy [3, 4].

In the last few decades, several data mining and machine-learning techniques have been developed for breast cancer detection and classification [5–7]. These approaches can be divided into three main stages: preprocessing, feature extraction, and classification.

In order to facilitate interpretation and benefit pursuit analysis, preprocessing of mammography helps to improve the visibility of peripheral areas and intensity distribution. For this reason, several methods have been reported [8, 9].

Feature extraction is an important step in breast cancer detection, where the main features help in discriminating benign tumors from the malignant ones. After this, several image properties, such as smoothness, coarseness, depth, and regularity, are extracted by means of segmentation [10]. Various transform-based texture analysis techniques are applied to convert the image into a new form using the spatial frequency properties of the pixel intensity variations. The common techniques are wavelet transform [11], fast Fourier transform (FFT) [12], Gabor transforms [13], and singular value decomposition (SVD) [14]. To reduce the dimensionality of the feature representation, principal component analysis (PCA) [15] can be applied.

Many works have attempted to automate the diagnosis of breast cancer based on machine-learning algorithms. Yap et al. [16] used three different methods of deep learning to detect the ultrasound of breast lesions: a patch-based LeNet, a U-Net, and a transfer learning approach with a pretrained FCN-AlexNet. Two different datasets are used. The first

dataset consists of 306 images (60 malignant and benign), and the second dataset consists of 163 images (53 malicious and 110 benign). The best *F*-measure results obtained by Yab were 91% and 89%, respectively. Qiao et al. [17] used the BI-RADS data system to improve diagnosis accuracy through ultrasound. In this work, the authors applied a genetic algorithm for the selection of features and the AdaBoost classifier to distinguish between benign and malignant tumors. Experiments were conducted on 138 tumors from the database using cross-checking of exit. The obtained accuracy was 93.48%. Wang et al. [18] applied the support vector machine (SVM) algorithm for breast cancer diagnosis in order to minimize the variability and increase diagnostic accuracy. In this work, receiver-operating characteristic curve ensemble (WAUCE) was used as a performance measure. The results of employing 12 different types of SVM were 97.89% for variation and 33.34% for accuracy. Amrane et al. [19] proposed two different classifiers (k-nearest neighbors (KNN) and Naïve Bayes (NB)) to diagnose breast cancer. The results showed that KNN achieved the highest accuracy of 97.51%, and the lowest accuracy of NB was 96.19%. Sun et al. [20] explored a multimodal deep neural network model by integrating multidimensional data (MDNNMD) to predict breast cancer. The obtained results show that there are still some issues in predicting the survival time of breast cancer with MDNNMD effectively. The proposed algorithm achieved an accuracy of 79.4%. The convolutional neural network improvement for breast cancer classification (CNNI-BCC) was applied to extract the relevant features from mammogram images [21]. The classification of the given features with CNNI-BCC achieved an accuracy of 90.50%.

Comparisons between neural networks and deep neural network algorithms with and without dimensional reduction techniques, and linear discrimination analysis were applied in [22] to classify the Wisconsin Diagnostic Breast Cancer (WDBC) dataset. The results showed that these algorithms performed well in terms of disease diagnosis and classification. The classification results proved the closeness of the accuracy of both previous models. In [23], the authors presented a comparison of three algorithms (Naïve Bayes, RBF network, and J48) on 683 breast cancer datasets with 10-fold cross-validation. The results showed that Naïve Bayes had the best accuracy of 97.36%, while RBF network and J48 achieved accuracies of 97.77% and 93.41%, respectively. Mondal et al. [24] explored the entropy method with four machine-learning algorithms to distinguish between normal tissues and breast cancer. The comparison of SVM, NB, KNN, and random forest (RF) indicated that SVM outperforms the others with an accuracy of 91.5%. Zhou et al. [25] exploited shear-wave elastography (SWE) data with a convolutional neural network (CNN) for breast cancer diagnosis using a database with 540 images, where 315 were malignant and 222 benign. The study achieved an accuracy of 95.8%, sensitivity of 96.2%, and specificity of 95.7%. In order to improve the accuracy of breast cancer diagnosis, Tamilvanan [26] applied the dimensionality reduction technique to five classifiers: multilayer perceptron, NB, radial basis function network, conjunctive rule, and

KNN. Besides precision, recall, F-measure, accuracy, and ROC, new metrics were used, such as balanced classification rate (BCR) and Matthews correlation coefficient (MCC). The experiments revealed that the NB algorithm had the highest accuracy of 82%. Tapak et al. [27] compared and analyzed eight classifiers to predict breast cancer survival and metastasis. The classifiers used were AdaBoost, Naïve Bayes (NB), least-square SVM (LSSVM), random forest (RF), SVM, linear discriminant analysis, Adabag, and logistic regression (LR). The proposed models were applied to 550 patients. The deduced results demonstrate outperformance of SVM over the other machine-learning methods. In [28], the authors proposed the CRISP-DM methodology to analyze the WDBC dataset. The achieved best accuracy was for the SVM.

Tahmassebi et al. [29] applied magnetic resonance methods for detecting breast cancer in women (average age: 46.5 years, range: 25–70 years). Eight classifiers were used to categorize features, including linear SVM, linear discrimination analysis, logistic regression, decision tree, adaptive enhancement, and enhanced gradient extreme (XGBoost). To reduce the dimension, the authors applied various features selection techniques: survival-free redundancy (RFS) and survival of the disease (DSS). The rating accuracy was assessed with the area under the receiver operating characteristics curve (AUC). The best result was found with an AUC of 92%.

The Gabor wavelet has been proposed to extract features from mammography images in [30]. Several methods were used to classify radiographic images: C5.0 tree, SVM, artificial neural networks, Tree Quest, and CHAID. The comparison of the previous classifiers models demonstrates that SVM outperforms the other with an accuracy of 96%. Another deep learning algorithm was applied by Arau et al. [31] to recognize benign and malignant abnormalities. In this work, the authors used the convolutional neural networks (CNNs) to extract features. The given model was applied to a database with four classes (normal tissue, benign lesion, in situ carcinoma, and invasive carcinoma) and two classes (carcinoma and noncarcinoma). The sensitivity of this method was 95.6%.

In [32], the authors explored the tissue morphology in hematoxylin and eosin (H&E) stained breast cancer tissue microarray (TMA) samples using a machine-learning algorithm. The presented experiments showed that the morphological patterns in breast cancer can be explored using unsupervised machine learning.

Sharma et al. [7] applied three machine-learning algorithms to predict breast cancer. The experimental results gave an accuracy training ranging from 93% to 97%.

In order to ameliorate breast cancer histopathological image classification accuracy, Wei et al. [33] applied a deep CNN named BiCNN. The contribution of the authors was to propose a new technique to extract the features from mammography images. Radiya-Dixit et al. [34] developed a combined model with active feature extraction (CAFE) based on logistic regression (LR) techniques. In this work, the authors compared the proposed model to five machine-learning models. The best AUC was 91.8%.

In [35], the authors proposed an assistant tool based on deep learning to detect the breast cancer metastasis in lymph nodes. The developed tool demonstrates its effectiveness to identify micrometastases. Table 1 summarizes selected literature on breast cancer diagnosis. It highlights the type of technique applied to select the features, the classifier model to distinguish between malignant and benign, the accuracy measure, and the database used to validate the proposed mode.

While the higher accuracy of the various methods applied for the breast cancer diagnosis, the use of Tabu search metaheuristic [43] to reduce the dimension features improves the performance of the applied algorithms. The present work compares five machine-learning techniques to distinguish between malignant and benign breast cancer. The techniques include KNN [44], Gaussian Naïve Bayes (GNB) [45], logistic regression (LR) [46], the extremely randomized tree (ET) [47], and adaptive boosting [48].

The remainder of this paper is organized as follows. In Section 2, the material and methods are explained. Section 3 summarizes the experimental results, and Section 4 presents the conclusions.

2. Materials and Methods

2.1. Datasets Used for Research. In this work, two databases were used to validate the proposed algorithms. In fact, those who work with models of medical field are aware that relying on just one data source can be problematic. In addition two different methods have been applied to extract the features for each dataset; in BIDMC-MGH dataset, the feature extraction is basically composed by nuclei segmentation and nuclei computation; however, in WDBC dataset, the features were extracted depending on nuclei computation. The WDBC is frequently used in the comparison of BIDMC-MGH.

The BIDMC-MGH dataset was created in collaboration with the two centers, MGH and BIMDC. In addition to this, the two partners proceeded based on standardized laboratory protocol and the same equipment in their study. The BIDMC-MGH dataset [5] is composed of 392 features based on shape, intensity, texture, and color. The dataset contains 116 breast biopsies from MGH and 51 breast biopsies from BIDMC diagnosed as DCIS or UDH. The first 116 samples are used for training and the other 51 samples for validation of the proposed models. The features of BIDMC-MGH were computed from the morphological and statistical features of the selection nuclear regions. Based on shape and measurement values, the morphological features were computed including perimeter, area, bounding rectangle fit ellipse, shape descriptors, and Feret's diameter. The statistical features computations depend on the intensity, the texture, and eight chosen colors. Using the statistical analysis, the mean, median, and standard deviation are computed for each feature per patient. All images are extracted from the following website: http://earlybreast.becklab.org/.

The features of WDBC are subtracted from digitized images of a fine needle aspirate of a breast mass (FNA), which describes features of the nucleus of the current image.

The Wisconsin Diagnostic Breast Cancer (WDBC) database is composed of 569 observations, where 357 are benign and 212 are malignant breast masses. The 30 descriptive features produced from the three statistic values were computed based on ten FNA geometric measurements for each cell nucleus: radius, texture, perimeter, area, smoothness, compactness, concavity, concave point, symmetry, and fractal dimension of each mass.

2.2. Feature Selection. The feature selection is one of the most important steps in designing the classifier model. The main objective of feature selection operation is to get the best representation of the data which finds a lower dimensional representation of data. Usually high-dimensional representation of the data leads to degeneration in the performance of the used method. By considering that the redundant data should be removed and only the relevant feature will be used, the performance of the useful model can be improved or can be maintained and can simplify the complexity the applied model. In this context, the proposed feature selection method was presented. In this work, the Tabu search method based on rough set theory was introduced. The selection of the appropriate features depends on four stages: neighborhood search, diversification, shaking, and elite reduct. This process selects the informative information in order to discriminate between the normal and cancerous tissues; in addition, the large number of features increases the computing complexity. So the classification process will run slowly. Moreover, accurate diagnosis depends strongly on appropriate features being selected.

2.2.1. Rough Set Theory Based Feature Reduction. Recently, rough set theory has proved its effectiveness to reduce the features of a given dataset [49]. Mainly, this theory is based on discernibility and attribute dependency [50] to evaluate the contribution of each attribute. In rough set theory, all the attributes are called the decision system which is composed of conditional attributes set C as input and decision attributes set D as output.

Assume $I = (U, A)$ an information system, where U is a nonempty set of finite objects called universe of discourse and A is a nonempty set of attributes such that, for every $a \in A$, $a: U \longrightarrow Va$, where Va is the value set of a.

With any subset $P \subseteq A$, there is an associated equivalence relation called P-indiscernibility relation defined as follows: $\text{IND}_I(P) = \{(x, y) \in U \times U / \forall a \in P, a(x) = a(y)\}$. In other words, the objects x and y are indiscernible from each other by attributes from P. The equivalence classes of the P-indiscernibility relation are designated by $[x]_P$.

Let $X \subseteq U$; the set of object X can be approximated based on P-lower and P-upper approximations denoted by $\underline{P}X$ and $\overline{P}X$, respectively, where $\underline{P}X = \{x/[x]_P \subseteq t\, X\}$ and $\overline{P}X = \{x/[x]_P \cap tXn = q\varnothing\}$.

Assume P and Q are equivalence relations on U; the positive region Q with respect to P is defined as $\text{POS}_P(Q) = \cup_{X \in U/Q} \underline{P}X$. Based on positive region, the degree of dependency of Q frm P is defined as

TABLE 1: Summary of machine-learning algorithms for breast cancer diagnosis.

Author	Feature	Algorithm	Accuracy (%)	Dataset
Asri et al. [36]	FNA	SVM	97.13	UCI
Ivančáková et al. [28]	FNA	SVM	97.66	WDBC
Mondal et al. [24]	Entropy	SVM	91.5	Gene Expression Omnibus
Ghasemzadeh et al. [30]	Gabor wavelet	SVM	96	Mammography (DDSM)
Ayoub Shaikh and Ali [37]	Wrapper subset eval	SVM	97	Breast Cancer Digital Repository (BCDR)
Wang et al. [18]	Full features	SVM	33.34	SEER
Mengjie Yu [38]	Concave points	SVM	99.77	UCI
Wei et al. [33]	BiCNN	CNN	97	BreaKHis
Bejnordi et al. [39]	Morphology	CNN	92	WSIs
Arau et al. [31]	Full features	CNN	83	Histology Dataset
Yap et al. [16]	FCN-alexnet	CNN	98 92	B&K Medical Panther 2002 and B&K Medical Hawk 2102 US systems
Ting et al. [21]	Wise	CNN	90.50	Digital Mammogram
Zhou et al. [25]	CNN	CNN	95.8	SWE data
Sun et al. [20]	mRMR	Deep neural network	18.7	METABRIC/MDNNMD
Kaur et al. [40]	CNN	MLP	86	Mini-MIAS
Joshi et al. [22]	Scaling	NN	96.47	WDBC
Radiya-Dixit et al. [34]	Computational method	LR	91.8	BIDMC-MGH
Tahmassebi et al. [29]	Volume distribution	LR	92	WDBC
Braman et al. [41]	Heterogeneity	LR	93	ISPY1-TRIAL
Maysanjaya et al. [42]	Wrapper	NB	99.27.	UCI
Chaurasia et al. [23]	FNA	NB	97.36	WDBC
Tamilvanan [26]	Dimensionality reduction	NB	82	WDBC
Qiao et al. [17]	BI-RADS	AdaBoost	93.48	138 pathologically proven breast tumors
Turkki et al. [32]	Morphological	KNN	95	FinProg
Amrane et al. [19]	FNA	KNN	97.51	WDBC

$$\gamma_P(Q) = \frac{\left| POS_P(Q) \right|}{|U|}, \tag{1}$$

where $|U|$ defines the cardinality of the set U.

The computing of the attribute reduction for a large dataset is very expensive due to the existence of a set of reduction attributions. As a result, an alternative tool is necessary to overcome this concern.

2.2.2. Tabu Search Based on Feature Reduction. Tabu search (TS) [43] is a heuristic method of local search used to solve complex and many large optimization problems. It is classified as a local search method with adaptive memory.

In fact, its main feature is to memorize solutions or information search processes visited during the search to explore the research space beyond local optimality. An adaptive memory algorithm can effectively generate a neighborhood solution from the current solution and accepts the best solution even if it does not improve the current solution. The memory contains a list of recently visited solutions and avoids cycling and falling back into permanence in the local optimum. This simple process then allows escape from the local optimum so other areas of the solution space can be visited. To summarize, we can say that, at each iteration, the process continues to explore the solution space even though it does not improve the solution.

The TS-based attribute reduction [50] is mainly designed using the four strategies.

Strategy 1: neighborhood search

The basic goal of this strategy is to explore the solution around the current solution, besides it avoids generating a solution recently used in the Tabu List (TL).

Assume yi is the trial solution of the neighbors of x, where $I \in [1, m]$, and m is the number of trial solutions. An example of binary representation of a trial solution x is shown in Figure 1. The component of each vector equals zero or one according to the presence of each attribute in trial vector.

The extreme trial solutions are the vectors that contain zero or one for all the positions. The updating of the trial solutions is based on applying mutation multipoints, where the number of positions is generated randomly.

Strategy 2: diversification

In order to explore the space solution widely, the diversification will be applied. The diverse solution can be composed of the attributes not invoked during

x_1	x_2		x_3	x_4	....	x_{n-3}	x_{n-2}	x_{n-1}	x_n
1	0		0	1	...	0	1	1	0

FIGURE 1: Sample of trial solution.

the generation of trial solutions. The attributes with a low appearance will be selected with a probability inversely proportional to that appearing in the generation of the trial solutions.

Strategy 3: shaking

The purpose of this strategy is to optimize the current best solution. This procedure begins by deleting the contained attributes of the best solution one by one without any reducing of the dependency degree of $\gamma_{\_}$ ($x$best).
In other words, the updating of $x$best is approved only if the value of its dependency degree increased or kept the same value after removing an attribute that composes the solution.

Strategy 4: elite reduct

This strategy aims to find the minimal set attributes. Basically, this process depends on the previous strategy. In other words, the set of attributes involved in the minimal reduced vector is composed of the intersection of the list of vectors formed by optimizing the best solution. Figure 2 depicts the Tabu search based feature selection.

2.3. Machine-Learning Techniques. Often, the selection of classifier is one of the challenges in fielding of machine learning for the real applications. The choice of the classifiers algorithms depends on many factors. Some parameters are the complexity, the accuracy, the types of features/labels, and the suitability for certain sizes and dimensions of datasets.

The benefits of applying the Naïve Bayesian classifier are that it is less sensitive to the outliers, performs with less parameter, and also can outperform more alternative classifiers for small sample sizes.

The KNN is a supervised classifier that works well with small-size dataset. It is robust to a noisy data and requires one hyperparameter. Moreover, this technique can be used for both classification and regression problems.

Logistic performs better on a dataset with a small size. The effectiveness of the technique is to require few computational resources and does not need any tuning. The major strength of AdaBoost technique is its insusceptibility to overfitting problem.

The Extremely Randomized Trees (ET) model is relative to data with small number of samples and works with a reduced computational complexity.

2.3.1. K-Nearest Neighbors Algorithm. KNN was proposed by Cover and Hart and is considered one of the most successful machine-learning models to solve both classification and regression problems. The KNN technique is based on feature similarity measures. The aim of this method is to assign weights to the contributions of neighbors by assigning nearer neighbors with more weight than the more distant ones. The weight can be found based on the distance between instances. In general, distance measures can be standard Euclidean distance, Hamming distance, Manhattan distance, and Minkowski distance.

2.3.2. Gaussian Naïve Bayes. The Gaussian Naïve Bayes (GNB) classifier is a probabilistic machine-learning model and one of the most successful algorithms for classification tasks in medical image analysis [51]. The key insight of the classifier is conditional probability to classify data. The GNB algorithm is an example of NB algorithm where the features have continuous values and follow a Gaussian distribution.

2.3.3. Logistic Regression. LR, which is a statistics-based machine-learning algorithm, is generally applied for binary classification problems (problems with two class values). It is an effective method of modeling a categorical outcome (binomial/multinomial) with one or more independent variables. Unlike linear regression, which is used to study relationships between two continuous (quantitative) variables, LR is used to ascertain a probability value that can be mapped to two or more discrete classes.

2.3.4. Extremely Randomized Trees Classifier (ET). The ET is similar to RF, based on extreme randomization of the tree construction algorithm and disabling pruning operator. Ultimately, the main differences to RF are that it makes a small number of randomly chosen split points and it uses the whole learning sample. Besides better generalization abilities, the main strength of the ET technique is the reduction of the computational complexity.

2.3.5. AdaBoost Classifier (AB). Recently, the adaptive boosting gets a great interest in machine-learning competitions. It is proposed by Freund and Schapire in 1996. The technique focuses on building a classifier composed of a number of weak classifiers using the following equation:

$$F(x) = \text{sign}\left(\sum_{i=1}^{n} w_i f_i(x)\right), \qquad (2)$$

where w_i is the weight of the weak classifiers f_i.

This technique is based on the low correlation between the classifiers which significantly improve the accuracy of weak learning algorithms. At each step, the weights of the training examples were computed for each used model. The classifiers with low performances were kept and combined to produce the final results.

2.4. Evaluation Parameters. Various metrics are used to evaluate machine-learning algorithms. In this study, the useful metrics are accuracy, sensitivity, precision, *F*-score, and AUC.
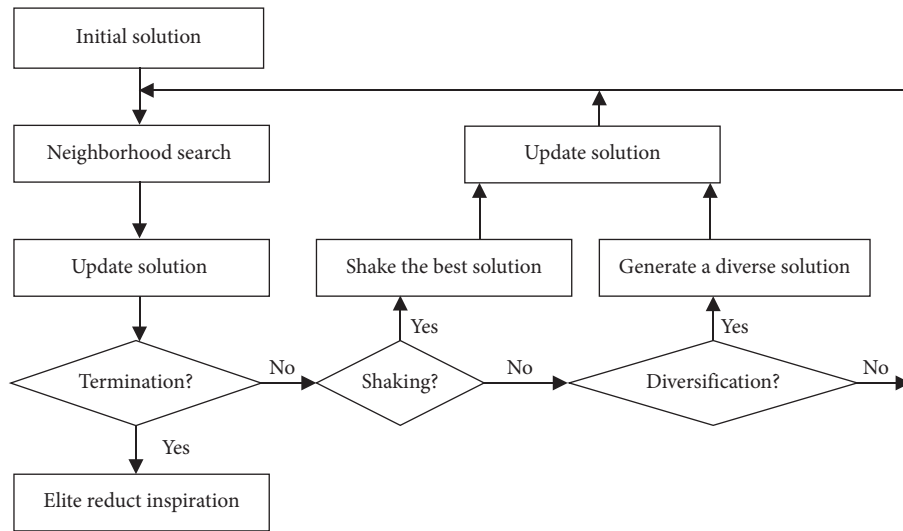
Figure 2: Feature selection strategy.

Accuracy is the measure of correct prediction of the classifier and provides general information about how many samples are misclassified. It is defined as

$$accuracy(\%) = \frac{TP + TN}{TN + FP + TP + FN} \times 100, \quad (3)$$

where TP, FP, TN, and FN are the numbers of true positives, false positives, true negatives, and false negatives, respectively, when the classifier is predicted.

Sensitivity is the ratio of the number of correctly predicted benign tumors to the total number of benign tumors:

$$sensitivity(\%) = TPR = \frac{TP}{TP + FN} \times 100. \quad (4)$$

Specificity is the proportion of actual malignant tumors that are classified as malignant by the model:

$$precision(\%) = \frac{TP}{TP + FP} \times 100,$$
$$F1score(\%) = \frac{precision \times sensitivity}{precision + sensitivity} \times 100. \quad (5)$$

In this study, besides the above metrics, the receiver-operating characteristic (ROC) graph AUC was employed.

3. Results and Discussion

To provide more evaluations, the proposed classifiers for breast cancer diagnosis were analyzed to study the effect of the TS technique on accuracy and to also compare it to techniques used in other works.

3.1. BIDMC-MGH Dataset. In this study, we compare the performance of five machine-learning techniques for the BIDMC-MGH database with 392 features and the WDBC database with 32 features. Several metrics were used to quantitatively evaluate the diagnostic performance of the classifiers. For the metrics described above, a higher

percentage indicates a better classification accuracy. Note that the AUC is a powerful metric for the classification of performance.

From the BIDMC-MGH dataset, 116 samples from the MGH hospital are used for training and 51 samples from the BIDMC hospital are used for testing. The TS was used to select features from all 392 features. Table 2 presents the comparison of the five classifiers using all features and without applying the TS feature selection. The best result to discriminate between malignant and benign cancer was related to using linear regression. As can be seen, all the applied classifiers for the BIDMC-MGH dataset did not exceed 83% in classification performance.

Table 3 lists the five metrics used for all classification techniques. As can be seen from Table 3, the AdaBoost presents the best models regarding accuracy, with an AUC score equal to 95%. As is known, AUC values provide a more accurate scoring measure than the other metrics based on true/false ratio. Figure 3 illustrates the comparison of five machine-learning algorithms for all the applied metrics.

Table 4 shows the comparison of the proposed machine-learning classifiers to other models using the same database. Performance evaluation is given by AUC ratio. The derived results show performance superiority of the AdaBoost classifier when compared to the models proposed in [5, 34]. Figure 4 shows the receiver-operating characteristic curves of the applied machine-learning algorithms in this experiment. The findings prove that Tabu feature selection can improve the accuracy of the applied classifiers. The feature selection process has an impact on most of the applied classifiers method. Figure 5 shows the effect of Tabu feature selection on the classifier accuracy. The TS technique reduces the number of features from 392 to 25. As a result, the obtained data become more understandable and easier to study.

3.2. WDBC Dataset. The proposed method was applied to the WDBC dataset. The effect of Tabu feature selection on

TABLE 2: Classifier results on BIDMC-MGH without Tabu feature selection.

Model	Accuracy (%)	Precision (%)	Sensitivity (%)	$F1$ score (%)	AUC (%)
KNN	72.54	75	73	73	74
GNB	68.62	70	69	69	69
LR	82.35	83	82	83	83
ET	72.54	72	73	72	69
AB	78.43	79	78	79	77

TABLE 3: Classifier results on BIDMC-MGH with Tabu feature selection.

Model	Accuracy (%)	Precision (%)	Sensitivity (%)	$F1$ score (%)	AUC (%)
KNN	74.50	79	75	74	76
GNB	78	79	76	77	78
LR	82.35	82	82	82	82
ET	90.19	91	90	90	89
AB	96.07	96	96	96	95

— k-Nearest neighbor classifier ROC (area = 0.76)
— AdaBoost classifier ROC (area = 0.95)
— Gaussian NB ROC (area = 0.78)
— Logistic regression ROC (area = 0.82)
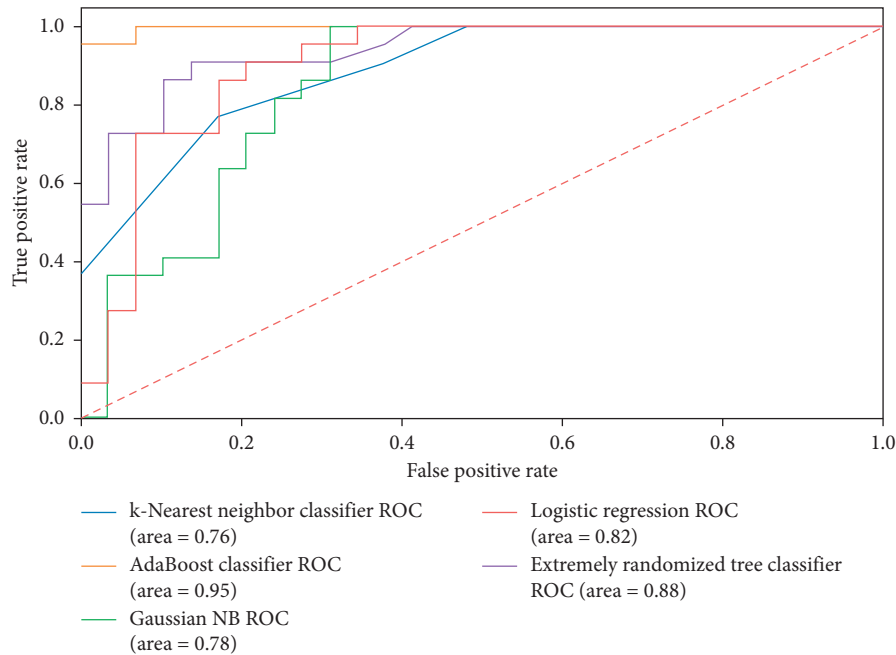— Extremely randomized tree classifier ROC (area = 0.88)

FIGURE 3: Analyzing the obtained results via different classifiers on BIDMC-MGH.

TABLE 4: Comparison of proposed method and other methods on BIDMC-MGH.

Model	AUC (%)
L1-regularized LR [5]	85.8
L1-regularized LR with active feature [5]	89.7
LR with early stopping and active features [34]	88.4
CAFE model [34]	91.8
Proposed model	**95**

accuracy is again analyzed in this section. The derived result will be compared to a list of machine-learning classifiers included in the review of the first section. The performance evaluation is measured by the accuracy metric for the comparison analysis because the methods presented in Table 1 used this evaluation criterion.

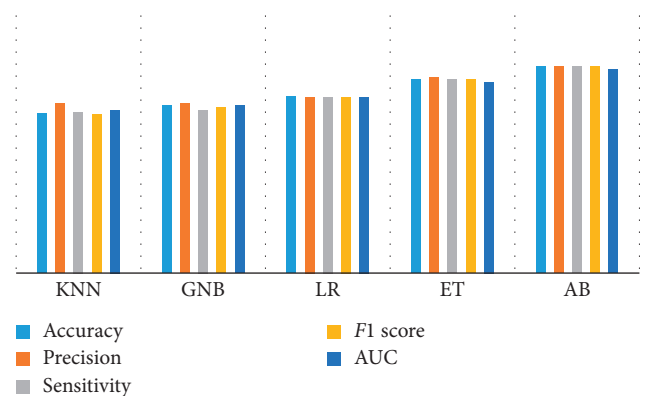■ Accuracy   ■ $F1$ score
■ Precision ■ AUC
■ Sensitivity

FIGURE 4: The ROC curve of the applied machine-learning algorithms.
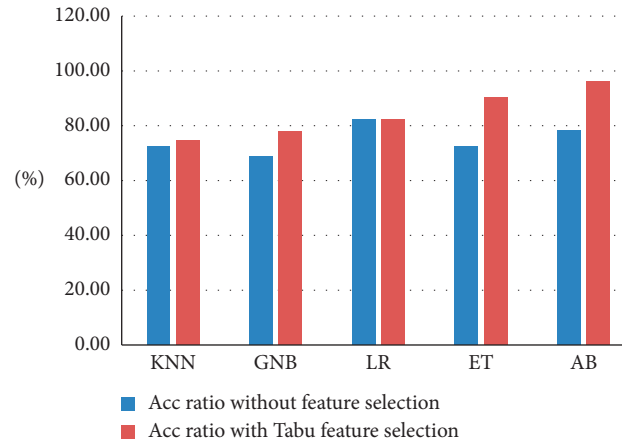
FIGURE 5: The effect of feature selection on accuracy on BIDMC-MGH.

TABLE 5: Classifier results on WDBC without Tabu feature selection.

Model	Accuracy (%)	Precision (%)	Sensitivity (%)	F1 score (%)	AUC (%)
KNN	97.07	97	97	97	97
GNB	92.98	93	93	93	93
LR	96.49	97	96	97	97
ET	95.32	96	95	95	96
AB	95.90	96	96	96	96

TABLE 6: Classifier results on WDBC with Tabu feature selection.

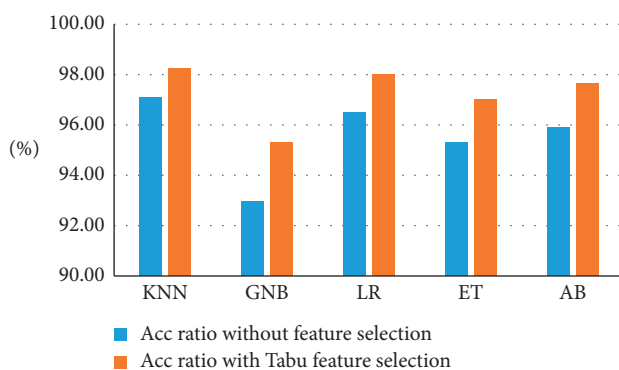| Model	Accuracy (%)	Precision (%)	Sensitivity (%)	F1 score (%)	AUC (%)
KNN	98.24	98	98	98	97
GNB	95.32	95	95	95	95
LR	98	99	99	99	98
ET	97	97	97	97	97
AB	97.66	98	98	98	97

FIGURE 6: Effect of feature selection on accuracy classifiers on WDBC.

In this experiment, the proposed classifiers were first applied to the WDBC dataset with all 32 features and then applied to the obtained TS-based features.

Table 5 illustrates the performance of the used classifiers before applying the TS method for feature selection, whereas Table 6 presents the evaluations performance of the machine-learning classifiers with Tabu features.

TABLE 7: Comparison of proposed method and other methods on WDBC.

Model	Accuracy (%)
NN [22]	96.47
KNN [19]	97.51
LR [29]	92
NB [23]	97.36
NB [26]	82
The proposed model	**98.24**

Figure 6 presents the effect of feature selection on the accuracy evaluation of the used classifiers. Table 7 shows the comparison of the proposed method with other methods in the literature. When we kept accuracy as the only criterion for evaluating performance, KNN was the best classifier model for classification of WDBC, whereas the linear regression method was the best classifier using the AUC metric.

4. Conclusions

Throughout this work, we developed an automated machine-learning technique for breast cancer diagnosis. The

proposed method was based on TS for the feature selection process, and five machine-learning algorithms were implemented to discriminate between malignant and benign cancer. The metrics for diagnosis of the BIDMC-MGH and WDBC datasets were evaluated using five evaluation criteria. Although the accuracy was used to evaluate the performance of the implemented models for breast cancer diagnosis, the AUC value along with the sensitivity, precision, and $F1$-score can in turn examine the evaluation process. In many cases, the derived conclusions based on accuracy metrics were totally different to those when using the AUC measure. We have shown via WDBC experiment comparisons that, in terms of AUC, KNN is the best classifier. However, this conclusion is not correct when considering accuracy. Moreover, the proposed classifiers based on the TS method demonstrate performance superiority over the other models. Future work includes extending the proposed algorithm for breast cancer to find the grade of malignant diagnosis and to apply the statistical methods to others pathology fields.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] J. Ferlay, I. Soerjomataram, R. Dikshit et al., "Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012," *International Journal of Cancer*, vol. 136, no. 5, pp. 359–386, 2014.

[2] J. G. Elmore, C. K. Wells, C. H. Lee, D. H. Howard, and A. R. Feinstein, "Variability in radiologists' interpretations of mammograms," *New England Journal of Medicine*, vol. 331, no. 22, pp. 1493–1499, 1994.

[3] W. H. Wolberg, W. N. Street, and O. L. Mangasarian, "Image analysis and machine learning applied to breast cancer diagnosis and prognosis," *Analytical and Quantitative Cytology and Histology*, vol. 17, no. 2, pp. 77–87, 1995.

[4] V. Meraner, J. Giesinger, G. Kemmler et al., "Development of a screening tool for the identification of psychooncological treatment need in breast cancer patients," *Psycho-Oncology*, vol. 18, no. 9, pp. 974–983, 2009.

[5] F. Dong, H. Irshad, E.-Y. Oh et al., "Computational pathology to discriminate benign from malignant intraductal proliferations of the breast," *PLoS One*, vol. 9, no. 12, Article ID e114885, 2014.

[6] B. Du, Q. Qi, Y. H. Zheng, X. Huang, and X. Ding, "Breast cancer histopathological image classification via deep active learning and confidence boosting," in *Artificial Neural Networks and Machine Learning - ICANN 2018*, vol. 11140, pp. 109–116, Springer, Berlin, Germany, 2018.

[7] A. Sharma, S. Kulshrestha, and S. Daniel, "Machine learning approaches for breast cancer diagnosis and prognosis," in *Proceedings of the International Conference on Soft Computing and its Engineering Applications*, Changa, India, December 2017.

[8] D. N. Ponraj, M. E. Jenifer, P. Poongodi, and J. S. Manoharan, "A survey on the preprocessing techniques of mammogram for the detection of breast cancer," *Journal of Emerging Trends in Computing and Information Sciences*, vol. 2, no. 12, pp. 651–664, 2011.

[9] A. P. Charate and S. B. Jamge, "The preprocessing methods of mammogram images for breast cancer detection," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 5, no. 1, pp. 261–264, 2017.

[10] P. Salembier and L. Garrido, "Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval," *IEEE Transactions on Image Processing*, vol. 9, no. 4, pp. 561–576, 2000.

[11] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechie, "Image coding using wavelet transform," *IEEE Transactions on Image Processing*, vol. 1, no. 2, pp. 205–220, 1992.

[12] G. Carter, C. Knapp, and A. Nuttall, "Estimation of the magnitude-squared coherence function via overlapped fast fourier transform processing," *IEEE Transactions on Audio and Electroacoustics*, vol. 21, no. 4, pp. 337–344, 1973.

[13] A. Teuner and B. J. Hosticka, "Adaptive gabor transformation for image processing," *IEEE Transactions on Image Processing*, vol. 2, no. 1, pp. 112–117, 1993.

[14] O. Edfors, M. Sandell, J.-J. van de Beek, S. K. Wilson, and P. O. Borjesson, "OFDM channel estimation by singular value decomposition," *IEEE Transactions on Communications*, vol. 46, no. 7, pp. 931–939, 1998.

[15] J. Yang, D. Zhang, A. F. Frangi, and J.-Y. Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 131–137, 2004.

[16] M. H. Yap, G. Pons, J. Martí et al., "Automated breast ultrasound lesions detection using convolutional neural networks," *IEEE Journal of Biomedical and Health Inform*, vol. 22, pp. 1218–1226, 2017.

[17] M. Qiao, Y. Hu, Y. Guo, Y. Wang, and J. Yu, "Breast tumor classification based on a computerized breast imaging reporting and data system feature system," *Journal of Ultrasound in Medicine*, vol. 37, no. 2, pp. 403–415, 2017.

[18] H. Wang, B. Zheng, S. W. Ko Yoona, and H. S. Sang Kob, "A support vector machine-based ensemble algorithm for breast cancer diagnosis," *European Journal of Operational Research*, vol. 267, no. 2, pp. 687–699, 2018.

[19] M. Amrane, S. Oukid, I. Gagaoua, and T. Ensari̇, "Breast cancer classification using machine learning," in *Proceedings of the 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, pp. 1–4, Istanbul, Turkey, April 2018.

[20] D. Sun, M. Wang, and A. Li, "A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 16, 2018.

[21] F. F. Ting, Y. J. Tan, and S. S. Kok, "Convolutional neural network improvement for breast cancer classification," *Expert Systems with Applications*, vol. 120, 2018.

[22] A. Joshi and A. Mehta, "Breast cancer classification using neural network and deep neural network techniques," *International Journal of Recent Scientific Research*, vol. 9, pp. 25788–25792, 2018.

[23] V. Chaurasia, S. Pal, and B. Tiwari, "Prediction of benign and malignant breast cancer using data mining techniques,"

Journal of Algorithms & Computational Technology, vol. 12, no. 2, pp. 119–126, 2018.

[24] M. Mondal, R. Semwal, U. Raj, I. Aier, and P. Kumar Varadwaj, "An entropy-based classification of breast cancerous genes using microarray data," *Neural Computing & Applications*, pp. 1–8, 2018.

[25] Y. Zhou, J. Xu, Q. Liu et al., "A radiomics approach with CNN for shear-wave elastography breast tumor classification," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 1935–1942, 2018.

[26] B. Tamilvanan, "An efficient classifications model for breast cancer prediction based on dimensionality reduction techniques," *International Journal of Advanced Research in Computer Science*, vol. 9, no. 1, pp. 448–455, 2018.

[27] L. Tapak, N. Shirmohammadi-Khorram, P. Aminic, B. Alafchi, O. Hamidi, and J. Poorolajal, "Prediction of survival and metastasis in breast cancer patients using machine learning classifiers," *Clinical Epidemiology and Global Health*, vol. 7, 2018.

[28] J. Ivančáková, F. Babič, and P. Butka, "Comparison of different machine learning methods on Wisconsin dataset," in *Proceedings of the 2018 IEEE 16th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, Kosice, Slovakia, February 2018.

[29] A. Tahmassebi, G. J. Wengert, T. H. Helbich et al., "Impact of machine learning with multiparametric magnetic resonance imaging of the breast for early prediction of response to neoadjuvant chemotherapy and survival outcomes in breast cancer patients," *Investigative Radiology*, vol. 54, no. 2, pp. 110–117, 2019.

[30] A. Ghasemzadeh, S. S. Azad, and E. Esmaeili, "Breast cancer detection based on Gabor-wavelet transform and machine learning methods," *International Journal of Machine Learning and Cybernetics*, vol. 10, pp. 1–10, 2018.

[31] T. Arau, G. Aresta, E. Castro et al., "Classification of breast cancer histology images using convolutional neural networks," *PLoS One*, vol. 12, no. 6, Article ID e0177544, 2017.

[32] R. Turkki, D. Bychkov, N. Linder, J. Isola, H. Joensuu, and J. Lundin, "Exploration of tissue morphologies in breast cancer samples using unsupervised machine learning," in *Proceedings of the AACR Annual Meeting*, Washington, DC, USA, 2017.

[33] B. Wei, Z. Han, X. He, and Y. Yin, "Deep learning model based breast cancer histopathological image classification," in *Proceedings of the IEEE International Conference on Cloud Computing and Big Data Analysis*, Chengdu, China, April 2017.

[34] E. Radiya-Dixit, D. Zhu, and A. H. Beck, "Automated classification of benign and malignant proliferative breast lesions," *Scientific Reports*, vol. 7, 2017.

[35] D. F. Steiner, R. MacDonald, Y. Liu et al., "Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer," *The American Journal of Surgical Pathology*, vol. 42, no. 12, pp. 1636–1646, 2018.

[36] H. Asri, H. Moatassime, H. Al Moatassime, and T. Noel, "Using machine learning algorithms for breast cancer risk prediction and diagnosis," *Procedia Computer Science*, vol. 83, pp. 1064–1069, 2016.

[37] T. Ayoub Shaikh and R. Ali, "Applying machine learning algorithms for early diagnosis and prediction of breast cancer risk," in *Proceedings of 2nd International Conference on Communication, Computing and Networking*, pp. 589–597, Springer, Berlin, Germany, 2018.

[38] B. S. Mengjie Yu, *Breast Cancer Prediction Using Machine Learning Algorithm*, The University of Texas at Austin, Austin, TX, USA, 2017.

[39] B. E. Bejnordi, J. Linz, B. Glass et al., "Deep learning—based assessment of tumor associated stroma for diagnostic breast cancer in histopathology images," in *Proceedings of the IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*Melbourne, Australia, April 2017.

[40] P. Kaur, G. Singh, and P. Kaur, "Improved cancer detection in mammogram images using automated deep learning technique. International," *Journal of Computer Sciences and Engineering*, vol. 6, 2018.

[41] N. Braman, K. Ravichandran, A. Janowczyk, J. Abraham, and A. Madabhushi, "Predicting neo-adjuvant chemotherapy response from pre-treatment breast MRI using machine learning and HER2 status," *Journal of Clinical Oncology*, vol. 36, no. 15, p. 582, 2018.

[42] I. M. D. Maysanjaya, I. M. A. Pradnyana, and I. M. Putrama, "Classification of breast cancer using wrapper and Naïve Bayes algorithms," *Journal of Physics: Conference Series*, vol. 1040, 2017.

[43] F. Glover, "Tabu search-part I," *ORSA Journal on Computing*, vol. 1, no. 3, pp. 190–206, 1989.

[44] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[45] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice-Hall, Upper Saddle River, NJ, USA, 2003.

[46] C. R. Boyd, M. A. Tolson, and W. S. Copes, "Evaluating trauma care: the TRISS method. Trauma score and the injury severity score," *The Journal of Trauma: Injury, Infection, and Critical Care*, vol. 27, no. 4, pp. 370–378, 1987.

[47] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, no. 1, pp. 3–42, 2006.

[48] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[49] L. H. Bac and N. A. Tuan, "Using rough set in feature selection and reduction in face recognition problem," in *Proceedings of the Pacific-asia Conference on Knowledge Discovery and Data Mining*, Hanoi, Vietnam, May 2005.

[50] A.-R. Hedar, J. Wang, and M. Fukushima, "Tabu search for attribute reduction in rough set theory," *Soft Computing*, vol. 12, no. 9, pp. 909–918, 2008.

[51] O. Maier, C. Schröder, N. D. Forkert, T. Martinetz, and H. Handels, "Classifiers for ischemic stroke lesion segmentation: a comparison study," *PLoS One*, vol. 10, no. 12, Article ID e0145118, 2015.