RESEARCH ARTICLE

Enhancing Early Breast Cancer Detection Through Advanced Data Analysis

MD. ATIQUR RAHMAN[1]**, MOHAMED HAMADA**[2]**, (Senior Member, IEEE),**
SHAYLA SHARMIN[3]**, TANZINA AFROZ RIMI**[3]**, (Graduate Student Member, IEEE),**
ATIA SANJIDA TALUKDER[3]**, NAFEES IMRAN**[4]**, KHADIJATUL KOBRA**[5]**,**
MD RIDWAN AHMED[5]**, MD RABBI**[5]**, MD. MAFIUL HASAN MATIN**[1]**, AND M. AMEER ALI**[6]

[1]Department of Computer Science and Engineering, East West University, Dhaka 1212, Bangladesh
[2]School of Computer Science and Engineering, The University of Aizu, Aizuwakamatsu 965-0006, Japan
[3]Department of Computer Science and Engineering, Daffodil International University, Dhaka 1216, Bangladesh
[4]Department of Information Technology and Management (ITM), Daffodil International University, Dhaka 1216, Bangladesh
[5]Department of Software Engineering, Daffodil International University, Dhaka 1216, Bangladesh
[6]Tesserent, Melbourne, VIC 3000, Australia

Corresponding author: Md. Atiqur Rahman (atick.rasel@gmail.com)

ABSTRACT In recent years, breast cancer, originating from breast tissue, has become one of the significant global health challenges for women worldwide, with early detection crucial for improved survival rates. Researchers have proposed numerous detection techniques, and recently, machine learning-based methods have gained considerable attention due to their reusability and speed. Despite various models proposed by researchers for breast cancer detection, there is an ongoing need for more accurate models. This study proposes an enhanced machine-learning approach for breast cancer detection using the Wisconsin Breast Cancer (Diagnostic) (WDBC) dataset. We applied several data preprocessing techniques, including hypothesis testing, feature engineering, scaling, and feature selection. We trained 14 classifiers by selecting the 13 most significant features using a gradient boosting regressor with Bonferroni correction. Our proposed eXtreme Gradient Boosting model demonstrated superior performance, achieving 99.12% accuracy, 0.9767 precision, 1.0 recall, 0.9861 specificity, and 0.9882 F1-score. These results surpass those of previous studies, underscoring the model's potential for early and accurate breast cancer diagnosis. Furthermore, evaluations based on training time and Kappa score indicate that our eXtreme Gradient Boosting model is faster and more reliable.

INDEX TERMS SMOTE, ANOVA test, ensembling feature engineering, Bonferroni correction, breast cancer, classification, machine learning.

I. INTRODUCTION

Breast cancer is the most common malignant tumor globally, posing a significant health threat, especially to women [2]. It is a term used to describe when breast cells grow abnormally beyond their usual boundaries of the breast [3]. Breast cancer is a diverse disease with three main types: sporadic, familial, and hereditary. Hereditary breast cancers are linked to high-penetrance gene mutations and often present earlier, while familial cases share similar characteristics [4]. Women encounter a significantly higher incidence of breast cancer compared to other forms of cancer. Breast cancer is the

most often diagnosed form of cancer and is responsible for most deaths due to cancer among female patients, surpassing lung cancer as the most common disease globally [5], [6]. Breast cancer has two types: invasive and non-invasive. Non-invasive breast cancer does not spread to surrounding breast tissue, while invasive breast cancer does. Among invasive breast cancers, there are two primary subtypes. Invasive Ductal Carcinoma (IDC) is the most common, originating in the milk ducts, and Invasive Lobular Carcinoma (ILC) is the second most common type, starting in the milk-producing glands (lobules) [7].

The Global Cancer Statistics 2022 report shows that breast cancer is the most prevalent form of cancer among the female population across 185 nations globally, accounting

The associate editor coordinating the review of this manuscript and approving it for publication was Yongming Li.

for 2,308,897 new cases, representing 11.6% of all newly diagnosed cancer cases, with an associated mortality rate of 6.9% [6], [8]. The American Cancer Society (ACS) reported that in 2022, 287,850 women and 2,710 men were affected by breast cancer, with 43,250 women and 530 men dying from the disease, indicating that the death rate is approximately 98.79% in women [9]. Projections by the World Health Organization (WHO) anticipate that global cancer cases will reach 19.3 million by 2025, though incidence rates vary significantly worldwide. Mortality rates are 17% higher in developing countries than in developed ones, despite a higher overall incidence rate of 88% in developed countries [10]. Also, the GLOBOCAN Cancer Tomorrow prediction tool forecasts a 46% increase in breast cancer incidence by 2040 [11]. In Indonesia, 40.3% of the population is affected by breast cancer, with 16.6% of those diagnosed dying from the disease. Risk factors such as an unhealthy diet, smoking, and alcohol consumption contribute to the increased likelihood of developing breast cancer [12].

Early diagnosis of breast cancer significantly improves survival chances and reduces long-term mortality rates, while advanced-stage diagnosis results in decreased survival rates. Detecting breast cancer at early stages can notably enhance treatment outcomes and prognosis. However, early detection remains challenging due to unusual features, missing values, and other hidden impediments. The complexity of breast cancer classification and diagnosis makes the process more difficult. A shortage of radiologists further exacerbates this challenge. Radiologists face significant pressure to manually analyze diagnostic reports and develop treatment plans, a process that is both time-consuming and prone to errors [13], [14], [15], [16].

Effective screening technologies are crucial for the early detection of breast cancer signs and symptoms [17], [18]. Traditionally, mammography [19], ultrasound [20], magnetic resonance imaging (MRI) [21], and pathology tests [22] have been the most common methods for screening and diagnosis. Recently, advanced classification methods, such as Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Extreme Gradient Boosting (XGBoost), have been employed. These methods enable precise and rapid detection to improve patient outcomes significantly. Machine learning (ML) techniques are transforming the biomedical and healthcare fields, playing a vital role in the early detection of breast cancer. The use of ML for molecular tumor classification has gained considerable attention and can enhance breast cancer detection [23], [24], [25], [26].

This study is motivated by the limitations of existing machine learning classifiers in detecting and classifying breast cancer. Timely and accurate detection of diseases remains a significant challenge in healthcare research [27], [28]. The primary aim of this research is to enhance the early detection rate. The main contributions of this study are as follows:

TABLE 1. Overview of existing research on breast cancer.

Ref	Publisher	Year	Best approach	Accuracy
[32]	IEEE	2020	Stacking classifier	97.2%
[33]	Sciencedirect	2021	MLP	98%
[34]	IEEE	2022	AdaBoost+ Bagging	84.3%
[35]	Elsevier	2023	Averaged-perceptron model	98.4%
[36]	IAES	2023	SVM	97.78%
[37]	IEEE	2024	ELRL-E	97.6%
[24]	Nature	2023	Decision Tree	83%
[38]	Wiley	2023	XGBoost	98%
[39]	Nature	2023	Decision Tree	98%
[17]	Elsevier	2023	Voting classifier	98.77%

1) Evaluation of various machine learning classifiers to determine the most effective method for breast cancer classification.
2) Demonstration of comprehensive data preparation steps to significantly improve classification performance, ensuring more accurate breast cancer prediction.
3) Comparison of the proposed model's performance with state-of-the-art methods, highlighting its superior effectiveness in this domain.

The paper is organized as follows: Section II covers the literature review. The methodology appears in Section III. Sections IV and V discuss the experimental results and comparative analysis. Finally, Section VI concludes the paper.

II. LITERATURE REVIEW

Although researchers aim to develop a more accurate machine-learning model for predicting breast cancer, finding the best model remains challenging. A comparative analysis involving several machine learning classifiers appeared in the paper [29], which showed that SVM achieved the highest accuracy of 97%. In [30], logistic regression attained an accuracy of 95% using texture classification and maximum perimeter. Milon et al. applied SVM, KNN, logistic regression, random forests, and ANN models on the Wisconsin Breast Cancer (Diagnostic) (WDBC) dataset, demonstrating that the ANN model achieved the highest accuracy of 98.57%, a precision of 97.82%, and an F1-score of 0.9890 [31]. The models that classified breast cancer best appear in Table 1.

On the Wisconsin Breast Cancer (Diagnostic) dataset, Uddin et al. [17] applied Principal component analysis (PCA) for dimensionality reduction. They then used eleven classifiers for classification: SVM, KNN, AdaBoost, Gradient Boosting, Voting Classifier, MLP, Nearest Cluster Classifier, Random Forests, Decision Tree, Naïve Bayes, and Logistic Regression. They achieved the highest accuracy of 98.77% using the Voting Classifier. Similarly, Birchha and Nigam applied the MinMax normalization technique and then used the Averaged-Perceptron model for classification using the same dataset, achieving an accuracy of 98.40% [35]. Additionally, Batool and Byun et al. utilized the same dataset

in their paper [37]. They applied correlation and mutual information to select the most crucial features to achieve the highest accuracy of 97.6%, with a precision of 0.964, a recall of 1.0, and an F1-score of 0.981 using their proposed model, ELRL-E. This voting classifier model combines four classifiers: Extra Trees, Light Gradient Boosting Machine, Ridge, and Linear Discriminant Analysis. Furthermore, Kadhim Ajlan et al. worked on the same dataset and demonstrated that the Extreme Learning Machine (ELM) algorithm achieved 92.06% accuracy, 80.25% precision, 96.60% recall, 87.56% F1-score, and 90.27% specificity [40].

In the study referenced as [39], feature selection from the SEER breast cancer dataset [41] used Variance Threshold and Principal Component Analysis. The classification algorithms employed were Naive Bayes, AdaBoost, Gradient Boosting, XGBoost, and Decision Tree. Among these, the Decision Tree achieved the highest accuracy of 98%, while AdaBoost attained the highest recall of 0.99. Hayum et al. worked on two datasets, DDSM and MIAS. They utilized a bilateral filter for denoising, contrast stretching for image enhancement, color histograms for feature extraction, and MCSO for feature selection. Subsequently, they developed a deep learning-based approach named Fuzzy Convolutional Neural Network (FCNN) for classifying breast cancer. Their model achieved accuracies of 95.02% and 98.95% for the DDSM and MIAS datasets, respectively [42].

Laghmati et al. [43] applied PCA for feature transformation and grid search for hyperparameter tuning, followed by applying seven ML classifiers on the Mammographic Mass and WBCD [44] datasets. They achieved 96% and 95.35% accuracies for the Mammographic Mass and WBCD datasets, respectively, using the XGBoost classifier. However, when they used an ensemble technique combining AdaBoost and the S-LR models, they achieved the highest accuracy of 97.37% for the WBCD dataset, although the accuracy for the Mammographic Mass dataset dropped by 2.63%.

Singh et al. applied three feature selection techniques: Gravitational Search Algorithm (GSA), Emperor Penguin Optimization (EPO), and a hybrid approach of GSA and EPO named hGSAEPO. They then applied six machine learning classifiers on the Wisconsin Breast Cancer (Diagnostic) dataset for classification. Among the three selection approaches, a classifier with the hGSAEPO feature selection technique demonstrated superior performance with an accuracy of 98.31%, sensitivity of 0.97, specificity of 0.9887, precision of 0.98, and an F1-score of 0.9539 [45].

Mahesh et al. proposed a blended ensemble approach combining five classifiers: SVM, KNN, decision tree, random forests, and logistic regression for breast cancer classification. Their model achieved an accuracy of 98.14% [46]. Sahu et al. [47] proposed a deep learning-based breast cancer detection by combining AlexNet, ResNet, and MobileNetV2. They applied a Laplacian of Gaussian-based modified high boosting filter (LoGMHBF) for data preprocessing. The model achieved an accuracy of 97.75% for

TABLE 2. Sample count by category in the initial dataset.

diagnosis	Sample Count
B	357
M	212

malignancy detection on the mini-DDSM dataset and 94.62% for malignancy detection on the BUSI dataset. Additionally, it provided an accuracy of 97.50% on the BUS2 dataset.

Gupta et al. [48] proposed a hybrid ensemble approach using an Artificial Neural Network and Deep Belief Network applied to the Breast Cancer Wisconsin dataset for classification. Their model achieved an accuracy of 98.14%. Baghel et al. proposed a Stacked Neural Network for breast cancer classification with the same dataset. They demonstrated accuracies of 98.51% in training and 97.2% in validation [49].

III. METHODOLOGY
A. DATASET DESCRIPTION
This research used the Wisconsin Breast Cancer (Diagnostic) dataset, available at the UC Irvine Machine Learning Repository [1]. This binary classification dataset contains 569 samples and 32 features. Of these, 31 features are numerical and independent, while one feature, 'diagnosis,' is categorical and dependent. The dependent feature has two categories: M and B, representing malignant and benign, respectively. The number of samples for each category appears in Table 2. Table 2 indicates that benign (B) accounts for 62.74% of the total samples, whereas malignant (M) accounts for 37.26%.

B. MISSING VALUE HANDLING AND CATEGORICAL ENCODING
The 'id' feature consists of unique values and does not provide significant information for the classification task, so we eliminated it at the start of the data preprocessing. We apply Equation (1) to determine whether the dataset (D) contains missing values. The results show values of zero for all features, indicating that it has no missing values. So, we applied the label encoder to encode the categorical feature 'diagnosis' [50]. The encoder assigned 0 and 1 for B and M, respectively.

$$MissingEntriesCount = D.isnull().sum() \qquad (1)$$

C. DATA SPLITTING
We then divided the dataset into training and test sets, with 80% of the samples allocated to the training set and the remaining 20% to the test set. The sample distribution in the training set appears in Table 3, where 285 samples belong to Category 0 and 170 samples belong to Category 1. This division represents 62.64% of samples from Category 0 and 37.36% from Category 1, indicating an imbalanced training set.

TABLE 3. Sample count by category in the training set.

diagnosis	Sample Count
0	285
1	170

TABLE 4. Sample count by category in the balanced training set.

diagnosis	Sample Count
0	570
1	455

D. BALANCING THE TRAINING SET

To mitigate the imbalance, we employed the SMOTE (Synthetic Minority Over-sampling Technique) method [51], [52], which generates new synthetic instances by interpolating between existing data points. Applying SMOTE almost balanced the training set, with 570 samples in Category 0 (55.61%) and 455 in Category 1 (44.39%). This balance helps reduce bias and improves the model's performance and reliability (Table 4).

E. DETERMINING THE STATISTICAL SIGNIFICANCE OF FEATURES

Subsequently, we conducted a one-way ANOVA test [53], [54] on the training set for hypothesis testing, setting a significance level of 0.05. Table 5 shows the p-value for each feature, with fractal_dimension_mean, texture_se, and symmetry_se having p-values of $9.22 \times e^{-01}$, $6.69 \times e^{-01}$, and $3.56 \times e^{-01}$, respectively, which exceeded the significance level. This result indicates insufficient evidence to reject the null hypothesis, implying a lack of significant association with the target. Consequently, we removed these features from the training and test sets from further analysis.

F. CREATING MEANINGFUL FEATURES

Improving the interpretability of a model allows for deeper insights into underlying data patterns and more accurate predictions. Moreover, you can enhance the interpretability by performing feature engineering, which generates new features from existing ones. Following hypothesis testing, we generated seven new features (created_feature1 to created_feature7) in the training and test sets separately using an ensembling technique [55] that combined seven models: Random Forest, Gradient Boosting, Gaussian Naïve Bayes, K-Nearest Neighbors (KNN), Decision Tree, Logistic Regression, and Support Vector Machine. We added these new features to the training and test sets, referring to them as engineered training and test sets. We also retained the original training and test sets as they were. The algorithm for generating the new seven features appears in Algorithm 1.

The training and test sets in the original feature space are indicated by O_{train} and O_{test}, respectively, while the training and test sets in the engineered feature space are represented by E_{train} and E_{test}, respectively. After adding seven new features to the engineering feature spaces, E_{train} and E_{test}

TABLE 5. Feature significance p-values.

Feature	p-value
fractal_dimension_mean	$9.22 \times e^{-01}$
texture_se	$6.69 \times e^{-01}$
symmetry_se	$3.56 \times e^{-01}$
fractal_dimension_se	$2.91 \times e^{-02}$
smoothness_se	$1.77 \times e^{-02}$
concavity_se	$6.23 \times e^{-15}$
compactness_se	$9.51 \times e^{-22}$
fractal_dimension_worst	$7.64 \times e^{-28}$
symmetry_mean	$8.28 \times e^{-33}$
smoothness_mean	$1.09 \times e^{-39}$
concave points_se	$1.10 \times e^{-43}$
texture_mean	$2.16 \times e^{-47}$
smoothness_worst	$5.10 \times e^{-49}$
symmetry_worst	$2.93 \times e^{-51}$
texture_worst	$2.90 \times e^{-61}$
area_se	$4.35 \times e^{-73}$
perimeter_se	$4.28 \times e^{-79}$
radius_se	$1.63 \times e^{-83}$
compactness_worst	$2.35 \times e^{-98}$
compactness_mean	$5.38 \times e^{-106}$
concavity_worst	$2.75 \times e^{-127}$
concavity_mean	$5.44 \times e^{-145}$
area_mean	$7.15 \times e^{-151}$
area_worst	$1.09 \times e^{-164}$
radius_mean	$1.71 \times e^{-170}$
perimeter_mean	$1.78 \times e^{-178}$
radius_worst	$1.17 \times e^{-204}$
concave points_mean	$2.90 \times e^{-205}$
perimeter_worst	$4.79 \times e^{-209}$
concave points_worst	$3.25 \times e^{-222}$

contain 34 features each, while in the original feature space, O_{train} and O_{test} each contain 27 features.

G. FEATURE STANDARDIZATION

Certain features can disproportionately affect a model's predictions, making it crucial to enhance its reliability. Therefore, we used a scaling technique called StandardScaler [56], which standardizes features by scaling them to have a mean of 0 and a standard deviation of 1. The StandardScaler is defined by Equation 2, where x and z denote the original and standardized values of the feature, respectively, and μ and σ represent the mean and standard deviation of the features, respectively.

$$z = \frac{x - \mu}{\sigma} \quad (2)$$

H. OPTIMIZING FEATURE SET FOR MODEL PERFORMANCE

Feature selection [57] addresses overfitting and enhances generalization. Selecting the most significant features allows a model to achieve good results with fewer features. Conversely, using many features may reduce a model's performance. In this work, we conducted three experiments on the engineered and original feature spaces, selecting the seven, nine, and thirteen most critical features based on scores calculated by the gradient boosting regressor using the Bonferroni correction technique [58], [59], [60], [61].

Algorithm 1 Feature Augmentation

Require: $O_{\text{train}} \in \mathbb{R}^{n \times m}$ (Original training set with n samples and m features)

Require: $O_{\text{test}} \in \mathbb{R}^{p \times m}$ (Original test set with p samples and m features)

Ensure: $E_{\text{train}} \in \mathbb{R}^{n \times (m+7)}$ (Augmented training set with n samples and $m + 7$ features)

Ensure: $E_{\text{test}} \in \mathbb{R}^{p \times (m+7)}$ (Augmented test set with p samples and $m + 7$ features)

1: Initialize classifiers C_i for $i = 1, 2, \ldots, 7$ (Seven classifiers)
2: **for** $i = 1$ to 7 **do**
3:     Train classifier C_i on O_{train}
4: **end for**
5: **for** $i = 1$ to 7 **do**
6:     $G_{i,\text{train}} \leftarrow C_i(O_{\text{train}})$ (Generate features for training set from classifier C_i)
7: **end for**
8: **for** $j = 1$ to 7 **do**
9:     Train classifier C_j on O_{test}
10: **end for**
11: **for** $j = 1$ to 7 **do**
12:     $H_{j,\text{test}} \leftarrow C_j(O_{\text{test}})$ (Generate features for test set from classifier C_j)
13: **end for**
14: $E_{\text{train}} \leftarrow \begin{bmatrix} O_{\text{train}} & G_{1,\text{train}} & G_{2,\text{train}} & \ldots & G_{7,\text{train}} \end{bmatrix}$
15: $E_{\text{test}} \leftarrow \begin{bmatrix} O_{\text{test}} & H_{1,\text{test}} & H_{2,\text{test}} & \ldots & H_{7,\text{test}} \end{bmatrix}$

The feature selection process is in Algorithm 2. Figures 1 and 2 demonstrate the gradient-boosting regressor-based Bonferroni correction scores for each feature in the original and engineered feature spaces, respectively.

Algorithm 2 Feature Selection for Original and Engineered Feature Spaces

---

1: **Input:** Original training set O_{train}, original test set O_{test}, engineered training set E_{train}, engineered test set E_{test}, number of selected features k
2: **Output:** Selected feature subsets for original and engineered spaces
3: **Step 1:** Apply feature selection on original training and test sets
4:     $F_{\text{original}} \leftarrow \text{BonferroniGBoost}(O_{\text{train}}, k)$
5:     $O'_{\text{train}} \leftarrow O_{\text{train}}[F_{\text{original}}]$
6:     $O'_{\text{test}} \leftarrow O_{\text{test}}[F_{\text{original}}]$
7: **Step 3:** Apply feature selection on engineered training and test set
8:     $F_{\text{engineered}} \leftarrow \text{BonferroniGBoost}(E_{\text{train}}, k)$
9:     $E'_{\text{train}} \leftarrow E_{\text{train}}[F_{\text{engineered}}]$
10:     $E'_{\text{test}} \leftarrow E_{\text{test}}[F_{\text{engineered}}]$
11: **Return:** $O'_{\text{train}}, O'_{\text{test}}, E'_{\text{train}}, E'_{\text{test}}$

In the first experiment, we selected the top seven features: (concave points_mean, area_se, concavity_worst, concave

FIGURE 1. Feature importance scores for original feature space.

points_worst, area_worst, perimeter_worst, and area_mean) from the original feature space, and (created_feature2, created_feature5, created_feature1, created_feature7, concave points_mean, smoothness_se, and area_se) from the engineered feature space. In the second and third experiments, we selected the top nine and thirteen features from the original and engineered feature spaces. Notably, the ranks of the top 13 features from the original feature space within the engineered feature space were 5, 7, 28, 29, 25, 24, 15, 11, 23, 22, 17, 26, and 18. These rankings reveal the relative significance and impact of the features in both feature spaces, offering vital insights into the feature selection process.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

This section discusses the experimental results obtained from the original and engineered feature spaces across the three experiments, with the evaluation metrics used to assess the models' effectiveness.

A. EVALUATION METRICS

Merely achieving a slightly higher accuracy does not necessarily indicate that the model is good. Therefore, we evaluated the classifiers using nine metrics: Accuracy (Acc), Precision (Pre), Recall (Rec), F1-score (F1), Specificity (Spe), Area Under the Curve (AUC), Training Time (TT), and Cohen's
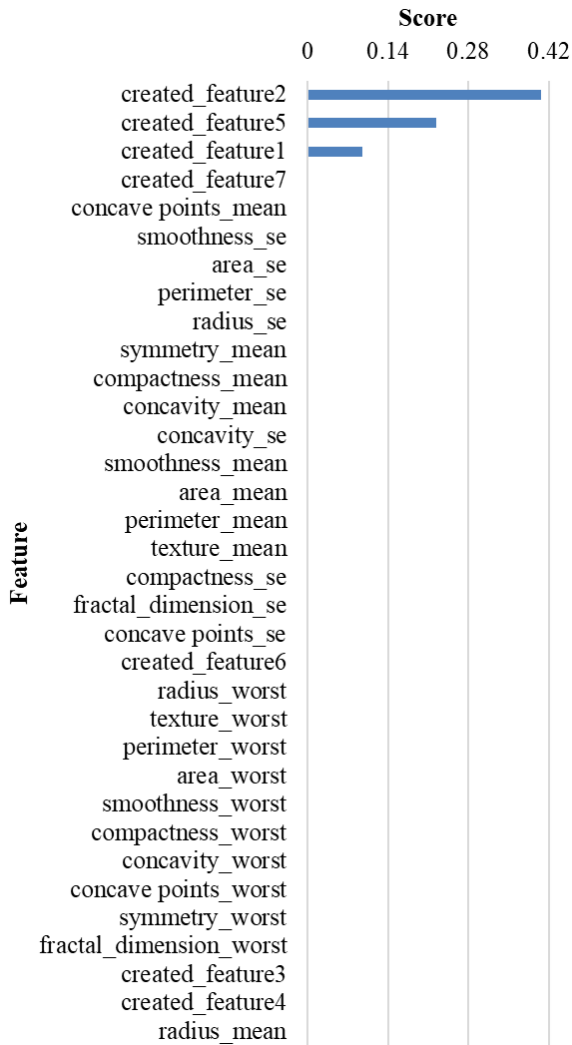
FIGURE 2. Feature importance scores for engineered feature space.

Kappa Score (Kappa) [62], [63], [64], [65], [66]. When evaluating a breast cancer classification model, various metrics comprehensively assess its performance. Accuracy measures the overall correctness of the model but may be misleading in cases of class imbalance. Precision indicates how many predicted positive cases are truly positive, which is crucial for minimizing false positives and avoiding unnecessary anxiety. Recall shows how well the model identifies all actual positive cases, which is critical for detecting every possible instance of breast cancer. F1-score balances precision and recall and offers a single metric when there is a trade-off between the two. Specificity assesses the model's ability to identify negative cases correctly and is essential for avoiding false alarms. Area Under the Curve (AUC) evaluates the model's discriminative power across various thresholds, providing insight into its overall performance. Training Time reflects the computational efficiency of the model and is critical for practical deployment and updates. Finally, Cohen's Kappa Score adjusts for class imbalance

and random chance, offering a more nuanced evaluation of the model's performance. Together, these metrics provide a holistic view of the model's effectiveness in breast cancer classification, ensuring that it performs well in identifying and differentiating between positive and negative cases. The metrics are formulated as follows, where TP, TN, FP, and FN represent True Positives, True Negatives, False Positives, and False Negatives, respectively.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{5}$$

$$\text{F1-score} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{6}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{7}$$

$$P_o = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$P_e = \frac{(TP+FP)(TP+FN)+(TN+FP)(TN+FN)}{(TP+TN+FP+FN)^2} \tag{9}$$

$$Kappa = \frac{P_o - P_e}{1 - P_e} \tag{10}$$

Here, P_o and P_e indicate the Observed Agreement and Expected Agreement.

B. COMPREHENSIVE CLASSIFICATION METRICS ANALYSIS
We applied 14 machine learning models for breast cancer classification on the original and engineered feature spaces, with 7, 9, and 13 features selected in three distinct experiments.

Table 6 shows the accuracy results. In the original feature space, the Light Gradient Boosting Machine and Extra Trees achieved 0.9825 accuracy with 7 features. With 9 features, the Light Gradient Boosting Machine, Tree-Based Pipeline Optimization Tool, and Extra Trees maintained this accuracy. When using 13 features, Extra Trees and eXtreme Gradient Boosting performed best, achieving 0.9912 accuracy in the original and engineered feature spaces. Thus, Extra Trees excelled with 13 features in the original space.

For the three experiments selecting 7, 9, and 13 features across both feature spaces, the precision, recall, F1-score, specificity, and AUC appear in Tables 7–9, respectively. Table 7 demonstrates that both Extra Trees and Light Gradient Boosting Machine achieved a precision of 1.0000, recall of 0.9524, F1-score of 0.9756, specificity of 1.0000, and AUC of 0.9762. These classifiers outperformed all other classifiers for these metrics in both feature spaces in experiment 1. Table 8 shows that the same two classifiers, Extra Trees and Light Gradient Boosting Machine, performed equally well when selecting nine features in the original feature space. However, the Tree-Based Pipeline

TABLE 6. Model performance: Accuracy in original vs. Engineered feature spaces.

	Accuracy for the Models					
Model Name	Selected Features (from Original Space)			Selected Features (from Engineered Space)		
	7	9	13	7	9	13
Support Vector Machine	0.9561	0.9474	0.9561	**0.9649**	0.9649	0.9649
Light Gradient Boosting Machine	**0.9825**	**0.9825**	0.9825	**0.9649**	**0.9737**	0.9737
eXtreme Gradient Boosting	0.9737	0.9649	0.9649	**0.9649**	**0.9737**	**0.9912**
Stacking Classifier	0.9649	0.9649	0.9737	0.9561	**0.9737**	0.9561
Voting Classifier	0.9649	0.9649	0.9737	0.9474	0.9561	0.9737
Decision tree	0.9211	0.9737	0.9298	0.9474	0.9474	0.9123
Catboost	0.9737	0.9737	0.9825	0.9474	0.9561	0.9737
Tree-Based Pipeline Optimization Tool	0.8947	**0.9825**	0.8947	0.9474	0.9386	0.9474
Gradient Boosting	0.9561	0.9649	0.9825	0.9474	0.9386	0.9737
Logistic Regression	0.9649	0.9474	0.9649	0.9474	0.9474	0.9649
Random Forest	0.9561	0.9474	0.9386	0.9386	0.9298	0.9386
Adaptive Boosting	0.9561	0.9649	0.9649	0.9386	0.9561	0.9825
Extra Trees	**0.9825**	**0.9825**	**0.9912**	0.9298	0.9386	0.9561
GridSearch CV-optimized model	0.9649	0.9561	0.9649	0.9035	0.9035	0.9035

Optimization Tool also matched the performance of these two classifiers while providing improvements of 8.33% in precision, 17.5% in recall, 13.26% in F1-score, 4.17% in specificity, and 10.67% in AUC in the original feature space. In contrast, the performance of Extra Trees, Light Gradient Boosting Machine, and Catboost remained stable across the same feature space. The Decision Tree performed 11.63% better in precision, 2.56% better in recall, 7.15% better in F1-score, 6.94% better in specificity, and 4.83% better in AUC in the original feature space while maintaining stability in the engineered feature space. Some models showed reduced performance for certain metrics; for instance, the voting classifier experienced declines in recall, F1-score, and AUC in the original feature space and precision and specificity in the engineered feature space. Table 9 reveals that eXtreme Gradient Boosting outperformed all other classifiers across the three experiments for the metrics of recall (1.0000), F1-score (0.9882), and AUC (0.9931) when applied to the engineered feature space with 13 features selected.

C. MODEL ASSESSMENT USING CONFUSION MATRIX AND ROC CURVE

Our study also evaluated the models using confusion matrices and ROC curves to assess their performance comprehensively. The confusion matrix is essential for identifying uneven class distribution and prevents misleading interpretations of classification accuracy. The ROC curve, on the other hand, illustrates a model's discriminative ability; a curve closer to the top-left corner indicates higher discriminative ability and better overall performance. Our results show that the Extra Trees and eXtreme Gradient Boosting classifiers achieved the highest accuracy of 0.9912 in the original and engineered feature spaces. Their confusion matrices and ROC curves appear in Figures 3-6.
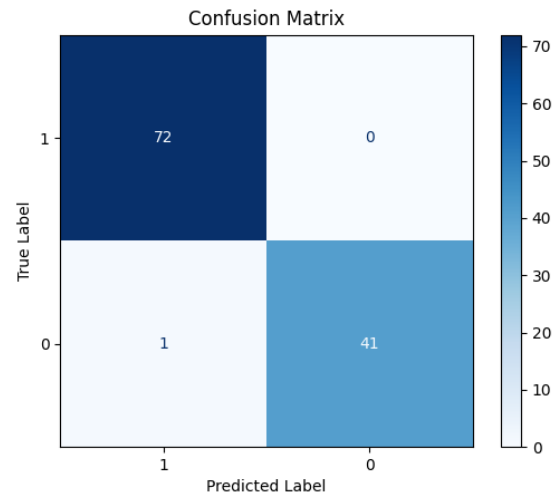

FIGURE 3. The confusion matrix of extra trees classifier.
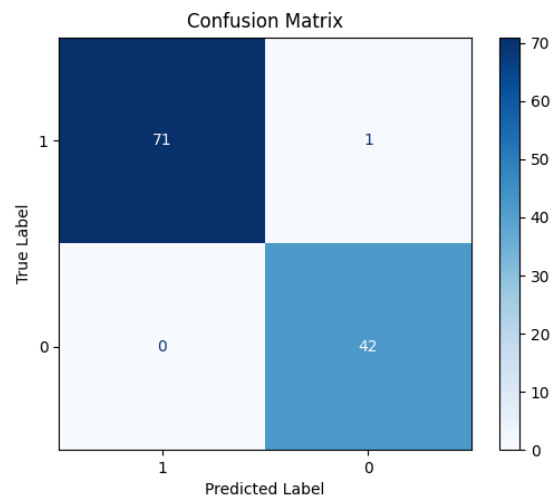

FIGURE 4. The confusion matrix of eXtreme gradient boosting.

Specifically, in the original feature space, Extra Trees misclassified class 1 as class 0 once in the original feature space, while eXtreme Gradient Boosting misclassified class 0 as class 1 once in the engineered feature space. Despite these misclassifications, the average AUC (Area Under the Curve) for Extra Trees was 0.9881, whereas for eXtreme Gradient Boosting was 0.9931. The ROC curve for eXtreme

TABLE 7. Performance metrics for models with 7 features.

Model Name	Seven Features (From Original Space)					Seven Features (From Engineered Space)				
	Pre	Rec	F1	Spe	AUC	Pre	Rec	F1	Spe	AUC
Extra Trees	**1.0000**	**0.9524**	**0.9756**	**1.0000**	**0.9762**	0.9048	0.9048	0.9048	0.9444	0.9246
Light Gradient Boosting Machine	**1.0000**	**0.9524**	**0.9756**	**1.0000**	**0.9762**	0.9524	**0.9524**	**0.9524**	0.9722	**0.9623**
Catboost	**1.0000**	0.9286	0.9630	**1.0000**	0.9643	0.9500	0.9048	0.9268	0.9722	0.9385
eXtreme Gradient Boosting	**1.0000**	0.9286	0.9630	**1.0000**	0.9643	**0.9750**	0.9286	0.9512	**0.9861**	0.9573
Voting Classifier	0.9750	0.9286	0.9512	0.9861	0.9573	0.9500	0.9048	0.9268	0.9722	0.9385
Stacking Classifier	0.9750	0.9286	0.9512	0.9861	0.9573	0.9512	0.9286	0.9398	0.9722	0.9504
GridSearchCV-optimized model	**1.0000**	0.9048	0.9500	**1.0000**	0.9524	0.9429	0.7857	0.8571	0.9722	0.8790
Logistic Regression	0.9750	0.9286	0.9512	0.9861	0.9573	0.9286	0.9286	0.9286	0.9583	0.9435
Random Forest	0.9744	0.9048	0.9383	0.9861	0.9454	0.9487	0.8810	0.9136	0.9722	0.9266
Adaptive Boosting	0.9744	0.9048	0.9383	0.9861	0.9454	0.9070	0.9286	0.9176	0.9444	0.9365
Gradient Boosting	0.9512	0.9286	0.9398	0.9722	0.9504	0.9500	0.9048	0.9268	0.9722	0.9385
Support Vector Machine	0.9512	0.9286	0.9398	0.9722	0.9504	**0.9750**	0.9286	0.9512	**0.9861**	0.9573
Decision Tree	0.8837	0.9048	0.8941	0.9306	0.9177	0.9286	0.9286	0.9286	0.9583	0.9435
Tree-Based Pipeline Optimization Tool	0.9167	0.7857	0.8462	0.9583	0.8720	0.9286	0.9286	0.9286	0.9583	0.9435

TABLE 8. Performance metrics with 9 features.

Model Name	Nine Features (From Original Space)					Nine Features (From Engineered Space)				
	Pre	Rec	F1	Spe	AUC	Pre	Rec	F1	Spe	AUC
Tree-Based Pipeline Optimization Tool	**1.0000**	**0.9524**	**0.9756**	**1.0000**	**0.9762**	0.9487	0.8810	0.9136	0.9722	0.9266
Light Gradient Boosting Machine	**1.0000**	**0.9524**	**0.9756**	**1.0000**	**0.9762**	0.9535	**0.9762**	**0.9647**	0.9722	**0.9742**
Extra Trees	**1.0000**	**0.9524**	**0.9756**	**1.0000**	**0.9762**	0.9268	0.9048	0.9157	0.9583	0.9315
Catboost	**1.0000**	0.9286	0.9630	**1.0000**	0.9643	0.9512	0.9286	0.9398	0.9722	0.9504
Decision Tree	**1.0000**	0.9286	0.9630	**1.0000**	0.9643	0.9286	0.9286	0.9286	0.9583	0.9435
eXtreme Gradient Boosting	**1.0000**	0.9048	0.9500	**1.0000**	0.9524	0.9535	**0.9762**	**0.9647**	0.9722	**0.9742**
Voting Classifier	**1.0000**	0.9048	0.9500	**1.0000**	0.9524	0.9302	0.9524	0.9412	0.9583	0.9554
Stacking Classifier	**1.0000**	0.9048	0.9500	**1.0000**	0.9524	**0.9756**	0.9524	0.9639	**0.9861**	0.9692
Adaptive Boosting	0.9750	0.9286	0.9512	0.9861	0.9573	0.9302	0.9524	0.9412	0.9583	0.9554
Gradient Boosting	**1.0000**	0.9048	0.9500	**1.0000**	0.9524	0.9070	0.9286	0.9176	0.9444	0.9365
GridSearchCV-optimized model	**1.0000**	0.8810	0.9367	**1.0000**	0.9405	0.9429	0.7857	0.8571	0.9722	0.8790
Random Forest	0.9500	0.9048	0.9268	0.9722	0.9385	0.9250	0.8810	0.9024	0.9583	0.9196
Logistic Regression	0.9500	0.9048	0.9268	0.9722	0.9385	0.9286	0.9286	0.9286	0.9583	0.9435
Support Vector Machine	**1.0000**	0.8571	0.9231	**1.0000**	0.9286	0.9750	0.9286	0.9512	**0.9861**	0.9573

TABLE 9. Model performance metrics with 13 features.

Model Name	Thirteen Features (From Original Space)					Thirteen Features (From Engineered Space)				
	Pre	Rec	F1	Spe	AUC	Pre	Rec	F1	Spe	AUC
Extra Trees	**1.0000**	**0.9762**	**0.9880**	**1.0000**	**0.9881**	0.9512	0.9286	0.9398	0.9722	0.9504
Catboost	**1.0000**	0.9524	0.9756	**1.0000**	0.9762	0.9756	0.9524	0.9639	**0.9861**	0.9692
Light Gradient Boosting Machine	**1.0000**	0.9524	0.9756	**1.0000**	0.9762	0.9333	**1.0000**	0.9655	0.9583	0.9792
Gradient Boosting	**1.0000**	0.9524	0.9756	**1.0000**	0.9762	0.9535	0.9762	0.9647	0.9722	0.9742
Voting Classifier	**1.0000**	0.9286	0.9630	**1.0000**	0.9643	0.9756	0.9524	0.9639	**0.9861**	0.9692
Stacking Classifier	**1.0000**	0.9286	0.9630	**1.0000**	0.9643	0.9302	0.9524	0.9412	0.9583	0.9554
eXtreme Gradient Boosting	**1.0000**	0.9048	0.9500	**1.0000**	0.9524	**0.9767**	**1.0000**	**0.9882**	**0.9861**	**0.9931**
GridSearchCV-optimized model	**1.0000**	0.9048	0.9500	**1.0000**	0.9524	0.9429	0.7857	0.8571	0.9722	0.8790
Adaptive Boosting	0.9750	0.9286	0.9512	0.9861	0.9573	0.9762	0.9762	0.9762	0.9861	0.9812
Logistic Regression	0.9750	0.9286	0.9512	0.9861	0.9573	0.9130	**1.0000**	0.9545	0.9444	0.9722
Support Vector Machine	0.9744	0.9048	0.9383	0.9861	0.9454	0.9524	0.9524	0.9524	0.9722	0.9623
Random Forest	0.9487	0.8810	0.9136	0.9722	0.9266	0.9268	0.9048	0.9157	0.9583	0.9315
Decision Tree	0.9474	0.8571	0.9000	0.9722	0.9147	0.8636	0.9048	0.8837	0.9167	0.9107
Tree-Based Pipeline Optimization Tool	0.9167	0.7857	0.8462	0.9583	0.8720	0.9500	0.9048	0.9268	0.9722	0.9385

Gradient Boosting is closer to the top-left corner than that of the Extra Trees classifier and all other classifiers evaluated in our experiments. This result indicates that eXtreme Gradient Boosting exhibits superior discriminative ability.

D. COMPUTATIONAL EFFICIENCY

We determined which model is the fastest by comparing the training times for the Extra Trees and eXtreme Gradient Boosting classifiers, as illustrated in Figure 7. Given that the
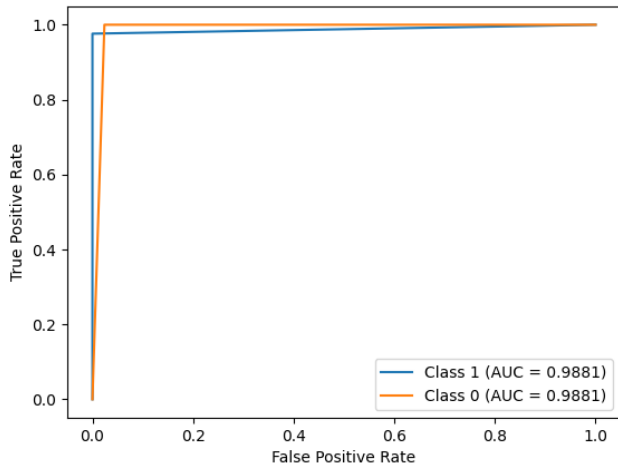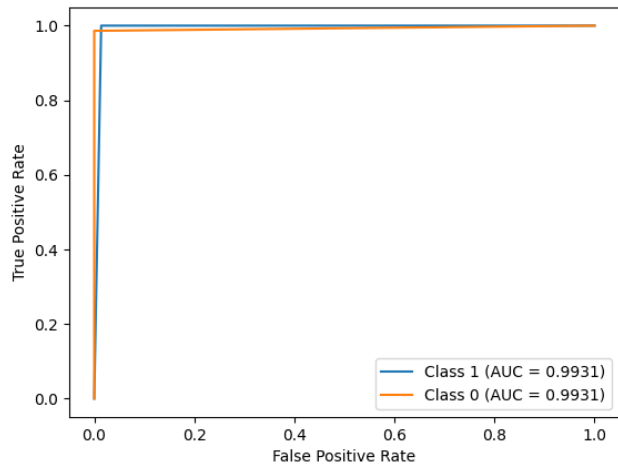
FIGURE 6. The ROC curve of eXtreme gradient boosting.

FIGURE 7. Comparison of training times.

TABLE 10. Cohen's kappa coefficients.

Model Name	Kappa (Original Space)	Kappa (Engineered Space)
eXtreme Gradient Boosting	Undefined	0.9779
Adaptive Boosting	-3.5752	-0.6738
Voting Classifier	Undefined	-2.1837
Catboost	Undefined	-2.1837
Optuna-optimized model	-8.7579	0.9509
Gradient Boosting	Undefined	0.1304
Support Vector Machine	-4.8187	-0.6272
Extra Trees	Undefined	-1.3177
Tree-Based Pipeline Optimization Tool	-2.5633	-1.9423
GridSearchCV-optimized model	Undefined	-4.3047
Light Gradient Boosting Machine	Undefined	0.928
Stacking Classifier	Undefined	-0.1211
Random Forest	-2.5169	-1.0022
Random Search-optimized model	-7.0115	0.9055
Logistic Regression	-3.5752	0.9055
Decision tree	-3.0354	-0.1025

training times were very short, we multiplied these times by 100 to clarify the comparison. The results indicate that eXtreme Gradient Boosting is 2.111 times faster than Extra Trees in the original feature space and 1.998 times more rapid in the engineered feature space.

E. COHEN'S KAPP ASSESSMENT
The Kappa statistic, also called Cohen's Kappa, assesses the agreement between the predicted and actual values beyond the possibility of the agreement occurring by chance and ranges from −1 to 1. A medical diagnosis requires a very high Kappa as it ensures the consistency of a model's prediction. The Kappa scores of all models on both feature spaces for experiment 3 appear in Table 10. The table shows that the Kappa scores for Extra Trees and Extreme Gradient Boosting in original and engineered feature spaces are Undefined and 0.9779, respectively. Extreme Gradient Boosting exhibits the highest Kappa among all models, which indicates its superior performance and suitability for breast cancer classification. A medical diagnosis uses precision to
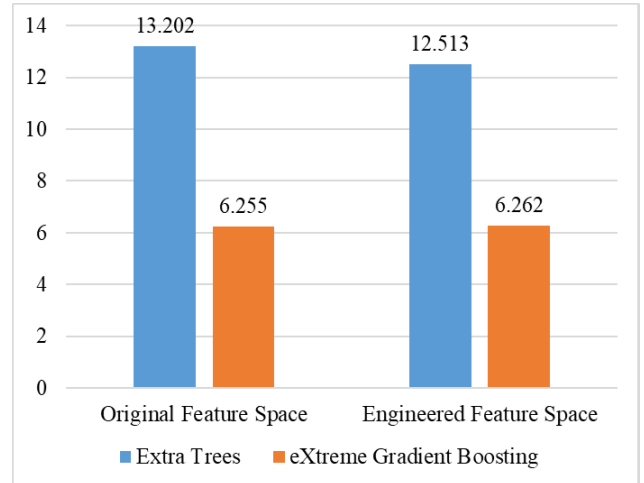
check false positives, meaning that incorrectly predicting breast cancer in this work. Lower recall suggests false negatives, indicating missing cases of cancer. Given the critical nature of breast cancer detection, maximizing recall is paramount to ensure that all cases are identified. Therefore, higher recall is more important than higher precision for breast cancer classification. As eXtreme Gradient Boosting demonstrated a perfect recall of 1.0 in the engineered feature space, which is higher than the recalls of all other models across all experiments, eXtreme Gradient Boosting is the best for breast cancer classification. The trained parameters of the recommended eXtreme Gradient Boosting model appear in Table 11.

F. REGULARIZATION AND OVERFITTING PREVENTION
To mitigate overfitting, we utilized L2 regularization with reg_lambda = 1 in our eXtreme Gradient Boosting model,

TABLE 11. Trained parameters of the recommended eXtreme gradient boosting model.

Parameter	Value
objective	binary:logistic
base_score	0.5
booster	gbtree
colsample_bylevel	1
colsample_bynode	1
colsample_bytree	1
enable_categorical	False
gamma	0
gpu_id	-1
grow_policy	depthwise
learning_rate	0.300000012
max_bin:	256
max_cat_threshold	64
max_cat_to_onehot	4
max_delta_step	0
max_depth	6
max_leaves	0
min_child_weight	1
n_estimators	100
num_parallel_tree	1
predictor	auto
reg_alpha	0
reg_lambda	1
sampling_method	uniform
scale_pos_weight	1
subsample	1
tree_method	exact
validate_parameters	1

TABLE 12. Performance summary.

Year [Ref]	Model (Dataset)	Instances and Selected Features	Accuracy	Precision & Recall	Specificity & F1-score
2024 [37]	ELRL-E (WDBC)	569 and 9	97.6%	- & 1.0	0.9643 & 0.9818
2023 [17]	Voting classifier (WDBC)	569 and 16	98.77%	0.9883 & 0.9854	- & 0.9868
2023 [39]	Decision Tree (SEER)	712319 and 13	98%	0.98 & 0.98	0.98 & 0.98
2023 [69]	t-SNE (WBCD)	699 and 9	86.6%	- & -	- & -
2024 [67]	SELF (WDBC)	569 and -	98.80%	0.9909 & 0.9909	- & 0.9909
2024 [40]	ELM (WDBC)	569 and -	92.06%	0.8025 & 0.9660	0.9027 & 0.8756
2024 [68]	OSELM (WBCD)	699 and -	97.89%	95.08% & 98.89%	97.41% & 96.93%
2024 [68]	OSELM (WDBC)	569 and -	96.13%	94.09% & 95.57%	96.51% & 94.80%
-	eXtreme Gradient Boosting (WDBC)	**569 and 13**	**99.12%**	**0.9767 & 1.0**	**0.9861 & 0.9882**

which penalizes large feature weights and helps to smooth the model, reducing its complexity. This regularization technique helps prevent the model from fitting noise in the training data and promotes better generalization to unseen data. Using L2 regularization ensures the model balances fit and complexity, supporting robust performance across various datasets.

V. PERFORMANCE COMPARISON

Comparative analysis is essential for evaluating a model's contribution to the field. By comparing the proposed model to existing benchmarks, we can assess its novelty and effectiveness. Table 12 presents a performance comparison of our proposed model with state-of-the-art techniques on Breast cancer datasets.

Previous studies, including [17], [37], [40], [67], [68], have reported varying performance metrics on the Wisconsin Breast Cancer (Diagnostic) dataset. In 2023, article [17] achieved an accuracy of 98.77%, a precision of 0.9883, a recall of 0.9854, and an F1-score of 0.9868 using 16 features, and published in a leading Elsevier journal but did not report specificity. As of 2024, articles [37], [40], [67], [68] also worked on the WBDC dataset. Among these, article [67] reported the highest accuracy of 98.80%, with precision, recall, and F1-score of 0.9909 each, but did not provide specificity or the number of features used. The three articles [40], [67], [68] were published in top Springer journals, while the article [37], published in IEEE Access 2024, achieved an accuracy of 97.6%. In contrast, our proposed model

demonstrates superior performance on the same dataset with an accuracy of 99.12%, a precision of 0.9767, a recall of 1.0, a specificity of 0.9861, and an F1-score of 0.9882, using only 13 features. This significant improvement demonstrates the effectiveness of our approach. Table 12 shows that our proposed method, eXtreme Gradient Boosting, outperformed all other methods mentioned. This significant improvement in performance metrics, coupled with a reduced number of features, underscores the effectiveness and efficiency of our approach, highlighting its potential impact and contribution to the field.

VI. CONCLUSION

This paper proposes an eXtreme Gradient Boosting classifier for detecting breast cancer, selecting only 13 features by a gradient boosting regressor with Bonferroni correction. Our model surpasses the state-of-the-art techniques with an accuracy of 99.12%, a precision of 0.9767, a recall of 1.0, a specificity of 0.9861, and an F1-score of 0.9882. Additionally, the experimental results indicate that our model demonstrated faster computational speed than other models. The findings of this study contribute to the advancement of breast cancer diagnosis and prognosis. Future research will focus on optimizing model parameters for improved performance.

REFERENCES
[1] O. M. W. Street, W. Wolberg. (2023). *Breast Cancer Wisconsin (Diagnostic)*. Accessed: Jul. 27, 2024. [Online]. Available: https://archive.ics.uci.edu/dataset/17/breast

[2] Y. Xu, M. Gong, Y. Wang, Y. Yang, S. Liu, and Q. Zeng, "Global trends and forecasts of breast cancer incidence and deaths," *Scientific Data*, vol. 10, no. 1, p. 334, May 2023.

[3] K. M. Cuthrell and N. Tzenios, "Breast cancer: Updated and deep insights," *Int. Res. J. Oncol.*, vol. 6, no. 1, pp. 104–118, 2023.

[4] A. Yiallourou, "Hereditary breast cancer syndromes breast cancer represents the most common type of cancer affecting the female population worldwide. According to the most recent data from the global cancer observatory (globocan), breast cancer has surpassed lung cancer as the leading cause of breast cancer management for surgeons: An examination guide," Medical School, Univ. Cyprus, Nicosia, Cyprus, Tech. Rep. 5, 2023, p. 79.

[5] W. Cao, K. Qin, F. Li, and W. Chen, "Comparative study of cancer profiles between 2020 and 2022 using global cancer statistics (GLOBOCAN)," *J. Nat. Cancer Center*, vol. 4, no. 2, pp. 128–134, Jun. 2024.

[6] F. Bray, M. Laversanne, H. Sung, J. Ferlay, R. L. Siegel, I. Soerjomataram, and A. Jemal, "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA, Cancer J. Clinicians*, vol. 74, no. 3, pp. 229–263, May 2024.

[7] B. He, H. Sun, M. Bao, H. Li, J. He, G. Tian, and B. Wang, "A cross-cohort computational framework to trace tumor tissue-of-origin based on RNA sequencing," *Sci. Rep.*, vol. 13, no. 1, p. 15356, Sep. 2023.

[8] D. S. M. Chan, L. Abar, M. Cariolou, N. Nanu, D. C. Greenwood, E. V. Bandera, A. McTiernan, and T. Norat, "World cancer research fund international: Continuous update project—Systematic literature review and meta-analysis of observational cohort studies on physical activity, sedentary behavior, adiposity, and weight change and breast cancer risk," *Cancer Causes Control*, vol. 30, no. 11, pp. 1183–1200, Nov. 2019.

[9] *Breast Cancer Death Rates Are Highest for Black Women-Again: Breast Cancer Facts & Figures*. Accessed: Jul. 26, 2024. [Online]. Available: https://www.cancer.org/research/acs-research-news/breast-cancer-death-rates-are-highest-for-black-women-again.html

[10] M. R. De Miglio and C. Mello-Thoms, *Reviews in Breast Cancer*. Switzerland: Frontiers in Oncology, Frontiers Media SA, 2023.

[11] E. Heer, A. Harper, N. Escandor, H. Sung, V. McCormack, and M. M. Fidler-Benaoudia, "Global burden and trends in premenopausal and postmenopausal breast cancer: A population-based study," *Lancet Global Health*, vol. 8, no. 8, pp. e1027–e1037, Aug. 2020.

[12] W. Wang, R. Jiang, N. Cui, Q. Li, F. Yuan, and Z. Xiao, "Semi-supervised vision transformer with adaptive token sampling for breast cancer classification," *Frontiers Pharmacol.*, vol. 13, pp. 62–74, Jul. 2022.

[13] Z.-R. Jiang, L.-H. Yang, L.-Z. Jin, L.-M. Yi, P.-P. Bing, J. Zhou, and J.-S. Yang, "Identification of novel cuproptosis-related lncRNA signatures to predict the prognosis and immune microenvironment of breast cancer patients," *Frontiers Oncol.*, vol. 12, Sep. 2022, Art. no. 988680.

[14] R. Ding, Y. Xiao, M. Mo, Y. Zheng, Y.-Z. Jiang, and Z.-M. Shao, "Breast cancer screening and early diagnosis in Chinese women," *Cancer Biol. Med.*, vol. 19, no. 4, pp. 450–467, Apr. 2022.

[15] M. Limb, "Shortages of radiology and oncology staff putting cancer patients at risk, college warns," *BMJ*, vol. 050011, p. 1430, Jun. 2022.

[16] K. Konstantinidis, "The shortage of radiographers: A global crisis in healthcare," *J. Med. Imag. Radiat. Sci.*, vol. 55, no. 4, Dec. 2024, Art. no. 101333.

[17] K. M. M. Uddin, N. Biswas, S. T. Rikta, and S. K. Dey, "Machine learning-based diagnosis of breast cancer utilizing feature optimization technique," *Comput. Methods Programs Biomed. Update*, vol. 3, Sep. 2023, Art. no. 100098.

[18] N. Routray, S. K. Rout, B. Sahu, S. K. Panda, and D. Godavarthi, "Ensemble learning with symbiotic organism search optimization algorithm for breast cancer classification and risk identification of other organs on histopathological images," *IEEE Access*, vol. 11, pp. 110544–110557, 2023.

[19] H. Yoen, M.-J. Jang, A. Yi, W. K. Moon, and J. M. Chang, "Artificial intelligence for breast cancer detection on mammography: Factors related to cancer detection," *Academic Radiol.*, vol. 31, no. 6, pp. 2239–2247, Jun. 2024.

[20] S. Balasubramaniam, Y. Velmurugan, D. Jaganathan, and S. Dhanasekaran, "A modified LeNet CNN for breast cancer diagnosis in ultrasound images," *Diagnostics*, vol. 13, no. 17, p. 2746, Aug. 2023.

[21] D. Sheth and M. L. Giger, "Artificial intelligence in the interpretation of breast cancer on MRI," *J. Magn. Reson. Imag.*, vol. 51, no. 5, pp. 1310–1324, May 2020.

[22] A. R. Vaka, B. Soni, and S. Reddy, "Breast cancer detection by leveraging machine learning," *ICT Exp.*, vol. 6, no. 4, pp. 320–324, Dec. 2020.

[23] R. K. Yadav, P. Singh, and P. Kashtriya, "Diagnosis of breast cancer using machine learning techniques—A survey," *Proc. Comput. Sci.*, vol. 218, pp. 1434–1443, 2023.

[24] M. Botlagunta, M. D. Botlagunta, M. B. Myneni, D. Lakshmi, A. Nayyar, J. S. Gullapalli, and M. A. Shah, "Classification and diagnostic prediction of breast cancer metastasis on clinical data using machine learning algorithms," *Sci. Rep.*, vol. 13, no. 1, p. 485, Jan. 2023.

[25] S. Hussain, M. Ali, U. Naseem, F. Nezhadmoghadam, M. A. Jatoi, T. A. Gulliver, and J. G. Tamez-Peña, "Breast cancer risk prediction using machine learning: A systematic review," *Frontiers Oncol.*, vol. 14, Mar. 2024, Art. no. 1343627.

[26] A. K. Das, S. K. Biswas, A. Mandal, A. Bhattacharya, and S. Sanyal, "Machine learning based intelligent system for breast cancer prediction (MLISBCP)," *Expert Syst. Appl.*, vol. 242, May 2024, Art. no. 122673.

[27] Imran, N. Iqbal, S. Ahmad, and D. H. Kim, "Health monitoring system for elderly patients using intelligent task mapping mechanism in closed loop healthcare environment," *Symmetry*, vol. 13, no. 2, p. 357, Feb. 2021.

[28] S. A. Yazdan, R. Ahmad, N. Iqbal, A. Rizwan, A. N. Khan, and D.-H. Kim, "An efficient multi-scale convolutional neural network based multi-class brain MRI classification for SaMD," *Tomography*, vol. 8, no. 4, pp. 1905–1927, Jul. 2022.

[29] V. Amudha, R. G. Babu, K. Arunkumar, and A. Karunakaran, "Machine learning-based performance comparison of breast cancer detection using support vector machine," in *Proc. AIP Conf.*, vol. 2676, 2022, Paper 050011.

[30] E. Halim, P. P. Halim, and M. Hebrard, "Artificial intelligent models for breast cancer early detection," in *Proc. Int. Conf. Inf. Manage. Technol. (ICIMTech)*, Sep. 2018, pp. 517–521.

[31] M. M. Islam, M. R. Haque, H. Iqbal, M. M. Hasan, M. Hasan, and M. N. Kabir, "Breast cancer prediction: A comparative study using machine learning techniques," *Social Netw. Comput. Sci.*, vol. 1, no. 5, pp. 1–14, Sep. 2020.

[32] M. R. Basunia, I. A. Pervin, M. Al Mahmud, S. Saha, and M. Arifuzzaman, "On predicting and analyzing breast cancer using data mining approach," in *Proc. IEEE Region 10 Symp. (TENSYMP)*, Jun. 2020, pp. 1257–1260.

[33] V. N. Gopal, F. Al-Turjman, R. Kumar, L. Anand, and M. Rajesh, "Feature selection and classification in breast cancer prediction using IoT and machine learning," *Measurement*, vol. 178, Jun. 2021, Art. no. 109442.

[34] F. Budiman, I. A. Saputro, P. Purwanto, and P. N. Andono, "Optimization of classification results by minimizing class imbalance on decision tree algorithm," in *Proc. Int. Seminar Mach. Learn., Optim., Data Sci. (ISMODE)*, Jan. 2022, pp. 6–11.

[35] V. Birchha and B. Nigam, "Performance analysis of averaged perceptron machine learning classifier for breast cancer detection," *Proc. Comput. Sci.*, vol. 218, pp. 2181–2190, 2023.

[36] N. Mohd Ali, R. Besar, and N. A. A. Aziz, "A case study of microarray breast cancer classification using machine learning algorithms with grid search cross validation," *Bull. Electr. Eng. Informat.*, vol. 12, no. 2, pp. 1047–1054, Apr. 2023.

[37] A. Batool and Y.-C. Byun, "Toward improving breast cancer classification using an adaptive voting ensemble learning algorithm," *IEEE Access*, vol. 12, pp. 12869–12882, 2024.

[38] H. Chen, N. Wang, X. Du, K. Mei, Y. Zhou, and G. Cai, "Classification prediction of breast cancer based on machine learning," *Comput. Intell. Neurosci.*, vol. 2023, no. 1, 2023, Art. no. 6530719.

[39] P. Manikandan, U. Durga, and C. Ponnuraja, "An integrative machine learning framework for classifying seer breast cancer," *Sci. Rep.*, vol. 13, no. 1, p. 5362, 2023.

[40] I. K. Ajlan, H. Murad, A. A. Salim, and A. F. B. Yousif, "Extreme learning machine algorithm for breast cancer diagnosis," *Multimedia Tools Appl.*, vol. 83, pp. 1–20, Jun. 2024.

[41] S. K. Mandala. (2023). *Seer Breast Cancer Data*. Accessed: Jul. 27, 2024. [Online]. Available: https://www.kaggle.com/datasets/ahmad03038/seer-breast-cancer-data-set

[42] A. A. Hayum, J. Jaya, R. Sivakumar, and B. Paulchamy, "An efficient breast cancer classification model using bilateral filtering and fuzzy convolutional neural network," *Sci. Rep.*, vol. 14, no. 1, p. 6290, Mar. 2024.

[43] S. Laghmati, S. Hamida, K. Hicham, B. Cherradi, and A. Tmiri, "An improved breast cancer disease prediction system using ML and PCA," *Multimedia Tools Appl.*, vol. 83, no. 11, pp. 33785–33821, Sep. 2023.

[44] D. Wolberg. *Breast Cancer Wisconsin (Original)*. Accessed: Jul. 27, 2024. [Online]. Available: https://archive.ics.uci.edu/dataset/15/breast

[45] L. K. Singh, M. Khanna, and R. Singh, "An enhanced soft-computing based strategy for efficient feature selection for timely breast cancer prediction: Wisconsin diagnostic breast cancer dataset case," *Multimedia Tools Appl.*, vol. 83, no. 31, pp. 76607–76672, Feb. 2024.

[46] T. R. Mahesh, V. V. Kumar, V. Vivek, K. M. K. Raghunath, and G. S. Madhuri, "Early predictive model for breast cancer classification using blended ensemble learning," *Int. J. Syst. Assurance Eng. Manage.*, vol. 15, no. 1, pp. 188–197, Jan. 2024.

[47] A. Sahu, P. K. Das, and S. Meher, "An efficient deep learning scheme to detect breast cancer using mammogram and ultrasound breast images," *Biomed. Signal Process. Control*, vol. 87, Jan. 2024, Art. no. 105377.

[48] V. Gupta, S. Wadhawan, V. Kumar, H. Sethi, and G. Gupta, "Breast cancer classification with ANN and DBN," in *Proc. 2nd Int. Conf. Advancement Comput. Comput. Technol. (InCACCT)*, vol. 2, May 2024, pp. 246–251.

[49] R. Baghel and E. Kumar, "Enhancing breast cancer diagnosis: Stacked neural networks for reducing data imbalance," in *Proc. 2nd Int. Conf. Disruptive Technol. (ICDT)*, vol. 132, Mar. 2024, pp. 1180–1183.

[50] T. Jo, *Machine Learning Foundations*. Cham, Switzerland: Springer, 2021, doi: 10.1007/978-3-030-65900-4.

[51] W. Banzhaf, P. Machado, and M. Zhang, *Handbook of Evolutionary Machine Learning*. Singapore: Springer, 2023.

[52] D. Elreedy, A. F. Atiya, and F. Kamalov, "A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning," *Mach. Learn.*, vol. 113, no. 7, pp. 4903–4923, Jul. 2024.

[53] K. Okoye and S. Hosseini, "Analysis of variance (ANOVA) in R: One-way and two-way ANOVA," in *R Programming: Statistical Data Analysis in Research*. Singapore: Springer, 2024, pp. 187–209.

[54] C. C. Aggarwal, *Probability and Statistics for Machine Learning: A Textbook*. USA: Springer, 2024.

[55] G. Kunapuli, *Ensemble Methods for Machine Learning*. New York, NY, USA: Simon and Schuster, 2023.

[56] B. Akshay, S. R. Pulari, T. Murugesh, and S. K. Vasudevan, *Machine Learning: A Comprehensive Beginner's Guide*. Boca Raton, FL, USA: CRC Press, 2024.

[57] A. Lindholm, N. Wahlström, F. Lindsten, and T. B. Schön, *Machine Learning: A First Course for Engineers and Scientists*. Cambridge, U.K.: Cambridge Univ. Press, 2022.

[58] P. Bahad and P. Saxena, "Study of adaboost and gradient boosting algorithms for predictive analytics," in *Proc. Int. Conf. Intell. Comput. Smart Commun. (ICSC)*. India: Springer, 2020, pp. 235–244.

[59] T. Bartz-Beielstein, S. Chandrasekaran, and F. Rehbach, "Case study II: Tuning of gradient boosting (XGboost)," in *Hyperparameter Tuning for Machine and Deep Learning With R: A Practical Guide*. Germany: Springer, 2023, pp. 221–234.

[60] G. Biau and B. Cadre, "Optimization by gradient boosting," in *Advances in Contemporary Statistics and Econometrics: Festschrift in Honor of Christine Thomas-Agnan*. France: Springer, 2021, pp. 23–44.

[61] F. Emmert-Streib, S. Moutari, and M. Dehmer, "Multiple testing corrections," in *Elements of Data Science, Machine Learning, and Artificial Intelligence Using R*. Germany: Springer, 2023, pp. 421–454.

[62] U. Braga-Neto, *Fundamentals of Pattern Recognition and Machine learning.*. Germany: Springer, 2020.

[63] Y. Deng, M. R. Eden, and S. Cremaschi, "Metrics for evaluating machine learning models prediction accuracy and uncertainty," *Comput. Aided Chem. Eng.*, vol. 52, pp. 1325–1330, 2023.

[64] P. Singh, *Fundamentals and Methods of Machine and Deep Learning: Algorithms, Tools, and Applications*. Hoboken, NJ, USA: Wiley, 2022.

[65] J. Tohka and M. van Gils, "Evaluation of machine learning algorithms for health and wellness applications: A tutorial," *Comput. Biol. Med.*, vol. 132, May 2021, Art. no. 104324.

[66] N. Fatima, L. Liu, S. Hong, and H. Ahmed, "Prediction of breast cancer, comparative review of machine learning techniques, and their analysis," *IEEE Access*, vol. 8, pp. 150360–150376, 2020.

[67] A. K. Jakhar, A. Gupta, and M. Singh, "SELF: A stacked-based ensemble learning framework for breast cancer classification," *Evol. Intell.*, vol. 17, no. 3, pp. 1341–1356, Jun. 2024.

[68] M. A. A. Albadr, F. T. Al-Dhief, L. Man, A. Arram, A. H. Abbas, and R. Z. Homod, "Online sequential extreme learning machine approach for breast cancer diagnosis," *Neural Comput. Appl.*, vol. 36, no. 18, pp. 10413–10429, Jun. 2024.

[69] N. Sharma, K. P. Sharma, M. Mangla, and R. Rani, "Breast cancer classification using snapshot ensemble deep learning model and t-distributed stochastic neighbor embedding," *Multimedia Tools Appl.*, vol. 82, no. 3, pp. 4011–4029, Jan. 2023.

MD. ATIQUR RAHMAN received the B.Sc. and M.Sc. degrees in computer science and engineering from the University of Rajshahi, Bangladesh, and the Ph.D. degree in computer and information systems from the School of Computer Science and Engineering, The University of Aizu, Japan. Currently, he is an Assistant Professor with the Department of Computer Science and Engineering, East West University, Bangladesh. He brings a wealth of expertise to his role. His strong skills in solving complex programming problems complement his robust publication record in international journals and conferences. In addition to his teaching and research responsibilities, he contributes to the academic community as a reviewer for various journals and conferences. His research interests include advancing the fields of machine learning, deep learning, and computer vision, reflecting his commitment to pushing the boundaries of these transformative technologies.

MOHAMED HAMADA (Senior Member, IEEE) received the Ph.D. degree from the University of Tsukuba, Japan, with a scholarship from the Japanese Government (MEXT). He is a Professor with The University of Aizu, Aizuwakamatsu, Fukushima, Japan. He has edited several books and authored over 100 papers in major international journals and conferences published by major publishers, such as IEEE, ACM, Elsevier, and Springer. He led several funded research projects and supervised graduate (M.Sc. and Ph.D.) and undergraduate students. His research interests include artificial intelligence and learning technologies. He is also interested in smart devices (such as smartphones and tablets) applications development and innovation. He is a member of the IEEE Technical Committee on Multimedia and the IEEE Technical Committee on Learning Technologies. He is a Senior Member of ACM Computer Society. He is a member of the editorial board of several international journals and a program committee member of several international conferences.

SHAYLA SHARMIN received the B.Sc. and master's degrees in CSE from JU Bangladesh. She is currently a Senior Lecturer with Daffodil International University. Before working in academia, she worked in various roles for leading ICT companies for more than four years. She has published many peer-reviewed academic papers in well-reputed conferences and journals in this area as an active Researcher. Her research interests include deep learning and image processing, particularly in medical imaging and healthcare analysis. As an Academician, she is an active member of several professional bodies.

TANZINA AFROZ RIMI (Graduate Student Member, IEEE) is a passionate Researcher in artificial intelligence and machine learning, focusing on medical image analysis and disease detection. Currently, she is with Daffodil International University. With a solid background in computer science and a keen interest in healthcare technology, she has dedicated her efforts to developing innovative solutions to address real-world challenges. Her goal is to leverage the power of AI to enhance healthcare outcomes and contribute to the betterment of society.

MD RIDWAN AHMED is currently pursuing the degree in software engineering with Daffodil International University in Bangladesh. Driven by a keen interest in the transformative potential of artificial intelligence, he is passionate about conducting research in AI, machine learning, and deep learning. His academic journey is characterized by a deep commitment to exploring and understanding the innovative applications of these technologies. At Daffodil International University, he engages in coursework and research that fuel his enthusiasm for advancing the field of artificial intelligence. He wants to leverage his knowledge and skills to contribute to significant advancements and impactful solutions within this dynamic and rapidly evolving domain.

ATIA SANJIDA TALUKDER has been a dedicated teacher working as a Lecturer with the Department of Computer Science and Engineering, since July 2023. She was a Lecturer with the Department of GED. During this period, she has taught various courses in statistics and probability, introduction to statistics, statistics for communication research with laboratory (using SPSS), and statistics for environment science. She is also a co-module leader and a module leader during this period. In addition, she is an active member of the committee to develop the OBE-based BSC syllabus for computational statistics under the Faculty of Science and Information Technology.

MD RABBI is currently pursuing the degree in software engineering with Daffodil International University, Bangladesh. His academic and research interests are deeply rooted in the transformative potential of artificial intelligence. Driven by a keen fascination with AI, he is passionate about researching machine learning and deep learning. As he advances through his studies, he is committed to exploring how these technologies can solve complex problems and drive innovation. His engagement with these cutting-edge fields reflects a dedication to understanding and contributing to the evolving landscape of technology. Daffodil International University equipped him with the skills and knowledge necessary to make meaningful strides in his research pursuits and to further the impact of AI in various applications.

NAFEES IMRAN received the B.B.A. degree from Bangladesh University of Business and Technology, Dhaka, Bangladesh, and the M.B.A. degree from Bangladesh University of Professionals. He is a Lecturer with the Department of Information Technology and Management (ITM), Daffodil International University, Dhaka, Bangladesh. His primary area of research is education. He has published several works in major conferences and journals. He has served as a reviewer for various international journals.

MD. MAFIUL HASAN MATIN received the bachelor's and master's degrees from Jahangirnagar University, Bangladesh. He is a Lecturer with the Department of Computer Science and Engineering, East West University, Bangladesh. With a strong passion for machine learning and deep learning, he dedicates his academic career to advancing knowledge in these fields. His work reflects a commitment to both teaching and research, contributing to the growing body of knowledge in computer science and its real-world applications.

KHADIJATUL KOBRA has been an innovative Lecturer of mathematics with Daffodil International University, for the past eight years. She has been teaching undergraduate students in several departments, including CSE, EEE, TE, SWE, and others. She is compassionate and dedicated to imparting crucial knowledge to students while fostering learning through various lesson plans and teaching strategies. Fuzzy mathematics is her area of study, which she enjoys applying to multiple research topics. She would like to use her research to benefit the world. She is a committed and enthusiastic researcher, who enjoys discovering new information and drawing connections between various subjects.

M. AMEER ALI received the Ph.D. degree in IT from Monash University. He is a seasoned IT professional with over 22 years of experience. He has authored more than 55 research publications and has held various roles in academia, including the department Head and the Dean of faculty. He was the Managing Director of IT company and a cyber security Consultant. His extensive expertise spans software development, network design, and cyber security, working with both private and public organizations. He holds several certifications, including PCI QSA, CEH, and ISO27001. Currently, he is a Senior Security Consultant, specializing in PCI and ISO27001 projects. His strong problem-solving skills, decision-making abilities, and proficiency in cyber security tools, make him a valuable asset to any organization.

● ● ●