ISSN 2355-3286

Feature Selection Benchmarks for Breast Cancer Diagnosis: A Comparative Machine Learning Study

R. Rossa Alfi Nur[1], Nashir Abbas Husaini[2], Moch. Arjunnaja[3], Az-Zahra Batrisyia Juniarto[4], Yuri Pamungkas[5*]
[1,2,3,4,5] Department of Medical Technology, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[1]5049231083@student.its.ac.id ; [2]5049231080@student.its.ac.id ; [3]5049231034@student.its.ac.id ; [4]5049231079@student.its.ac.id ; [5*]yuri@its.ac.id
* Corresponding Author

Accepted on February 13, 2026
Approved on June 03, 2026

Abstract—Breast cancer remains one of the most common causes of death among women, making early and precise detection essential. Yet conventional diagnosis can be limited by specialist shortages, cost, and slow workflows. We therefore assess machine-learning classification with feature selection to streamline diagnosis. Our contribution is a comparative benchmark of feature-selection strategies and classifiers on the WDBC dataset. We evaluated five models (SVM, neural-networks, decision tree, bagged-tree, and boosted-tree). Chi-square (χ^2), mRMR, and ReliefF selected 5, 10, 15, and 30 features, and performance was measured across multiple train–test splits using accuracy, precision, recall, specificity, and F1-score. SVM was overall the top performer and stable across splits. The best SVM setting reached 97.81% accuracy, with strong precision and F1-score, indicating reliable benign–malignant separation. Neural-networks usually ranked second but were more sensitive to the split. Bagged Tree generally improved on a single decision tree, while Boosted Tree showed mixed gains depending on the subset. ReliefF and mRMR often matched or exceeded χ^2 with smaller subsets, showing that careful feature reduction can retain accuracy while lowering dimensionality. In conclusion, combining effective feature selection with an appropriate classifier improves breast cancer classification, and SVM with a compact feature set is a practical choice.

Index Terms—Breast cancer classification; Machine learning; Feature selection; WDBC dataset; Support Vector Machine

I. INTRODUCTION

Cancer has become one of the most critical global health challenges today. Different forms of cancer, including lung, cervical, thyroid, and breast cancer, have generated widespread concern worldwide. Among women in particular, breast cancer stands out as one of the most common and life-threatening malignancies. [1]. Breast cancer is the most frequently diagnosed cancer among women, with concerning rates of occurrence and death. It is estimated that around 10% of women will develop breast cancer at some stage in their lifetime [2]. In the ranking of cancer-related mortality among women, breast cancer occupies the second position after lung cancer. This indicates that, despite not being the primary cause of cancer deaths, breast cancer remains a substantial contributor to female cancer mortality and represents a significant global public health burden [3]. Statistically, breast cancer contributes to 25% of all cancer cases and represents 12% of total new cancer cases diagnosed in women [4]. Medically, breast cancer is defined as uncontrolled cell proliferation originating from breast tissue [5]. This abnormal proliferation forms a mass called a tumor. Tumor classification is divided into two main categories: benign tumors which are non-cancerous, and malignant tumors which are cancerous [6]. Benign tumors have characteristics of slow growth, clear boundaries, and no ability to spread to other tissues or organs. In contrast, malignant tumors exhibit rapid and uncontrolled growth, possess poorly defined margins, and have the capacity to metastasize, spreading to distant organs or tissues throughout the body [7].

World Health Organization (WHO) data shows that breast cancer trends are experiencing significant increases not only in developed countries but also in middle- and low-income countries. Data indicate that more than 2.3 million females received a breast cancer diagnosis in 2020, with the number of deaths estimated at approximately 685,000 cases worldwide. WHO projections show an alarming increase, where the number of deaths from breast cancer increased from 685,000 to 963,000 in 2021 [8]. Breast cancer detection can be performed through various medical examination modalities, including breast ultrasonography, physical examination by professional medical personnel, biopsy (tissue sampling for histopathological analysis), mammography, and breast MRI [9]. Biopsy is considered the gold standard among all diagnostic techniques for breast cancer diagnosis confirmation. Radiologists and specialist doctors then interpret

comprehensive examination results to determine cell malignancy status [10].

However, there are several significant limitations in conventional diagnostic systems. First, limited availability of specialist doctors becomes a barrier to healthcare access. Second, most tumors in early stages do not show clear clinical manifestations. Third, the entire diagnostic process requires relatively long time. Fourth, high medical examination costs become a barrier for some communities [11]. Yet, early breast cancer detection can prevent patients from unnecessary therapy and significantly improve prognosis. The limitations in conventional diagnostic systems drive researchers to develop Computer-Aided Diagnosis (CAD) systems capable of detecting tumors and providing accurate and rapid results without dependence on doctor or radiologist interpretation [12]. Machine learning, as a branch of AI, has been widely applied for disease prediction based on historical data and learning patterns. This technology implements various approaches for performance optimization, including optimization, statistical, and probabilistic techniques.

Comparative studies have demonstrated that experienced specialist physicians achieve a diagnostic accuracy of approximately 79% in detecting breast cancer, whereas machine learning based approaches have been reported to attain accuracy levels of up to 91%, indicating their potential to enhance diagnostic performance [13]. These results show that machine learning has superior capabilities in preventing and detecting diseases compared to conventional methods. Thus, machine learning implementation has great potential to reduce mortality rates from cancer. In recent years, different approaches for forecasting breast cancer have emerged. Classification techniques such as Random Forest (RF), Support Vector Machine (SVM), AdaBoost Classifier, K-Nearest Neighbors (KNN), and XGBoost Classifier have been used in various recent literature. This study employs the Wisconsin Diagnostic Breast Cancer (WDBC) obtained from Kaggle and applies multiple machine learning based classification algorithms to systematically classify breast cancer types in patients with suspected malignancy. Five classification models used in this study include Decision Tree (DT), SVM, Neural Networks, Bagged Tree, and Boosted Tree. This research also implements three different feature selection methods, namely Chi-square (χ^2), Minimum Redundancy Maximum Relevance (MRMR), and ReliefF to optimize model performance.

This research has several main objectives: to conduct experimental analysis on the WDBC dataset from the UCI machine learning repository and examine the correlation between features and the observed target class, to implement various current machine learning classification models on the dataset with various feature selection methods, and to conduct comparative analysis of the results to identify the optimal combination between feature selection methods and classification algorithms. Based on the research background and objectives, this research attempts to answer two main research questions (primary research questions):

1) Among the three feature selection techniques (Chi-square (χ^2), MRMR, and ReliefF), which method demonstrates the highest effectiveness in enhancing the predictive accuracy of breast cancer classification?
2) Among the five classification algorithms (SVM, Neural Networks, DT, Bagged Tree, and Boosted Tree), which model exhibits superior performance based on comprehensive evaluation metrics, including accuracy, precision, specificity, recall, and F1-score?

The WDBC dataset has been used in many studies, but most of them test only one feature selection method or one classifier at a time, which makes it hard to tell which combination works best. This study differs by comparing several feature selection methods and classifiers together under the same settings, so that the most accurate classifier and the most efficient feature selection method can be identified, along with the trade-off between the number of features and accuracy.

II. Literature Review

Previous studies have applied various machine learning techniques to breast cancer detection using the WDBC dataset and its extensions. As summarized in Table I, recent literature highlights a clear progression in optimizing these models through feature selection and algorithm tuning. In 2021, Ara et al. [14] evaluated 30 features from the WDBC dataset, showing that the SVM achieved 96.5% accuracy while DT reached 95.1%. Similarly, Naji et al. [18] demonstrated that reducing the dataset to 11 features still yielded high accuracies of 98.4% for SVM and 98.8% for DT.

In 2022, studies began emphasizing advanced ensembles and optimizing pipelines. Aamir et al. [16] reported that supervised learning models, particularly multilayer perceptron (MLP), could achieve a very high diagnostic accuracy of 99.12% on the WDBC dataset. Furthermore, Rasool et al. [20] proved that optimized machine learning pipelines significantly improve breast cancer prediction, with SVM reaching 99.3% accuracy when combined with appropriate feature selection and Preprocessing.

TABLE I. Summary of Recent Work

Authors	Dataset	Algorithm	Best Accuracy
Ara et al. (2021) [14]	WDBC, 357 benign and 212 malignant, 30 features	SVM, DT	Acc: 96.5% (SVM), 95.1% (DT)
Naji et al. (2021) [18]	WDBC, 357 benign and 212 malignant, 11 features	SVM, DT	Acc: 98.4% (SVM), 98.8% (DT)
Aamir et al (2022) [16]	WDBC, 569 instances, 30 features	RF, GB, SVM, ANN, MLP	Acc: 99.12% (MLP)

Authors	Dataset	Algorithm	Best Accuracy
Rasool et al (2022) [20]	WDBC, 569 instances, 30 features	SVM, LR, KNN, EC	Acc: 99.3% (SVM)
Uddin et al. (2023) [13]	WDBC, 357 benign and 212 malignant, 30 Features	SVM, DT	Acc: 98.07% (SVM), 94.20% (DT)
Khalid et al. (2023) [15]	WDBC, 11 features	DT	Acc: 93.8% (DT)
Strelcenia and Prakoonwit. (2023) [17]	Wisconsin Breast Cancer Diagnosis Dataset	LR, RF, DT, KNN, MLP, XGBoost	Acc: 98.64% (DT)
Al-Imran et al. (2024) [19]	WDBC, 699 instances - NM	SVM	Acc: 98.5% (SVM)
Arifin et al. (2024) [22]	WDBC from Metabric, 15 features	SVM, DT	Acc: 72.9% (SVM), 90.6% (DT)
Okundalaye and Özdemir (2026) [21]	WDBC	SVM, RF, k-NN	Acc: 98.0% (SVM)

Subsequent research in 2023 continued to underscore the critical role of feature engineering and algorithm selection. Uddin et al. [13] and Khalid et al. [15] consistently affirmed the reliability of SVM and DT algorithms, obtaining Accuracy ranging from 93.8% to 98.07% depending on the chosen feature subsets. Moreover, Strelcenia and Prakoonwit [17] showed that feature engineering remains highly effective, with DT reaching 98.64% accuracy in a comparative evaluation of six classifiers.

More recently, studies by Al-Imran et al. [19] and Arifin et al. [22] in 2024 expanded evaluations across different configurations, maintaining SVM's position as a robust classifier with accuracies up to 98.5%. Finally, Okundalaye and Özdemir [21] further confirmed that optimized machine learning workflows based on WDBC can maintain strong predictive performance, with SVM outperforming RF and k-NN under rigorous feature selection and cross-validation.

Taken together, these studies indicate that breast cancer classification performance depends not only on the choice of classifier, but also heavily on feature engineering, dimensionality reduction, and validation strategies.

III. Methodology

The proposed approach to enhance predictive accuracy in breast cancer diagnosis involves the implementation of multiple machine learning algorithms, including DT, SVM, Neural Networks, Bagged Tree, and Boosted Tree, to generate classification predictions. With 5 parameters, 10 parameters, 15 parameters, and 30 parameters tested, resulting in TP, TN, FP, and FN. Figure 1 presents a summary of the research stages in the form of a flow diagram.

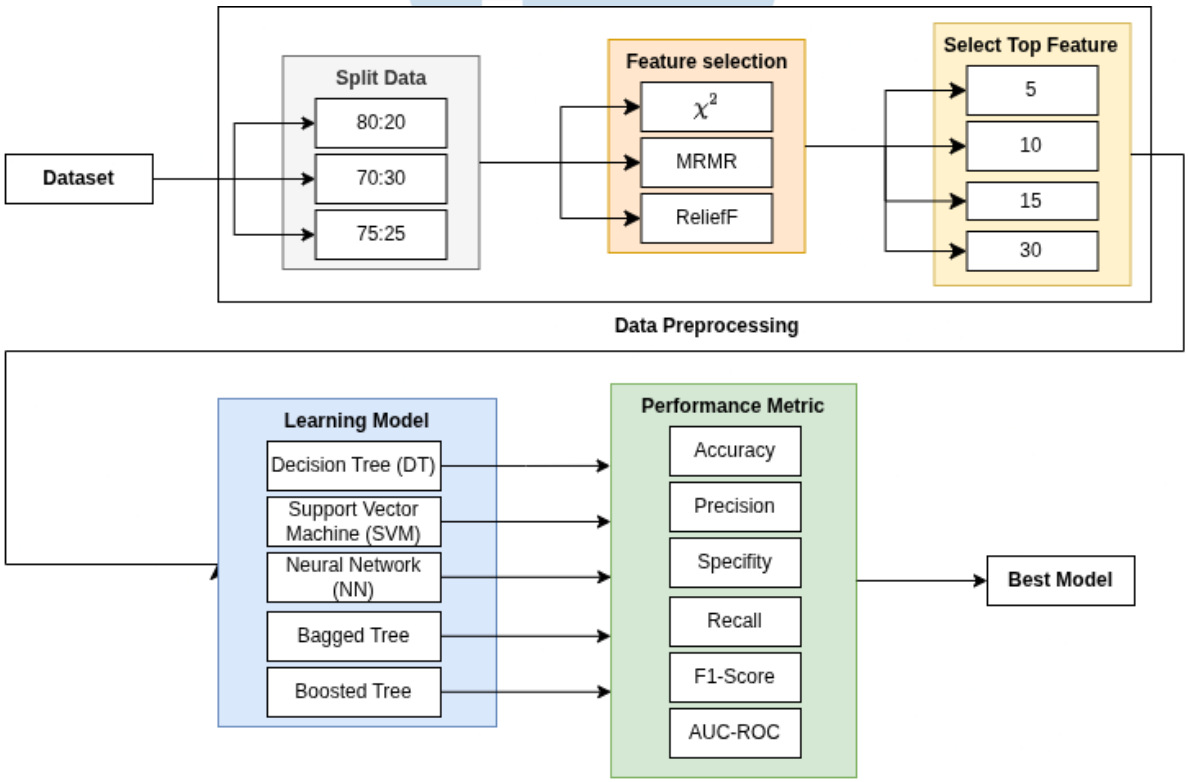

Fig. 1. Methodology of the research

A. Breast Cancer Dataset

This research uses a public dataset available on Kaggle named Wisconsin Diagnostic Breast Cancer (WDBC). This dataset contains parameters used to distinguish between benign (non-tumorous) and malignant (cancer-related) conditions. The dataset consists of 569 data instances and 30 numeric features

obtained from the UCI ML Repository, and contains no missing values. These features are not raw images, they are numeric values computed from digitized images of fine-needle aspirate (FNA) samples of breast masses, where each image was processed to measure characteristics of the cell nuclei (such as radius, texture, perimeter, and area). The classification in this study is therefore performed on these 30 computed features rather than on image pixels. From the total data, the dataset includes 357 benign cases and 212 malignant cases. The description of the 30 features is presented in Table II.

TABLE II. DESCRIPTIONS OF FEATURES

Feature Name	Feature Description
Radius	Mean distance from the centroid to points located along the lesion boundary.
Texture	Variation in grayscale intensity values across the region of interest.
Perimeter	Total length of the boundary contour surrounding the lesion.
Area	Total number of pixels enclosed within the lesion boundary, including a weighted contribution from boundary pixels.
Smoothness	Measure of local variation in radial distances, computed from differences between adjacent radial lengths.
Compactness	Ratio of the squared perimeter to the area, reflecting the degree of shape compactness.
Concavity	Quantitative assessment of the depth or prominence of concave regions along the contour.
Concave points	Number of discrete concave segments identified along the lesion boundary.
Symmetry	Degree of similarity between two halves of the lesion when divided along the major axis, measured by differences in perpendicular distances to the boundary.
Fractal Dimension	Estimated fractal dimension of the contour, representing boundary complexity; higher values indicate increased irregularity, which may correlate with greater malignancy risk.

B. Data Splitting

To ensure robust model evaluation, the dataset was partitioned using stratified random sampling with three training-to-testing ratios: 80:20, 75:25, and 70:30. The use of multiple train–test ratios was intended to assess the sensitivity of classification performance to the amount of training data available. Stratified sampling was applied to preserve the original class distribution of benign and malignant cases in both the training and testing subsets.

For each train–test configuration, a 5-fold cross-validation procedure was performed on the training dataset. The training data were divided into five approximately equal folds. In each iteration, four folds were used to train the model, while the remaining fold was used for validation. This process was repeated five times so that each fold served once as the validation set. The final cross-validation performance was obtained by averaging the results across all five folds. After model validation, the optimized model was evaluated on the independent test set corresponding to the selected train–test ratio.

C. Feature Selection

In this section, the dataset undergoes ranking iterations considering the top 5, 10, 15, and 30 features, where this ranking uses several feature selection algorithms: χ^2, MRMR, ReliefF.

1) Chi-square (χ^2)

Chi-square is a statistical approach that assesses how categorical variables are associated with the target class [23]. The χ^2 method is utilized in this research, as multi-class data is supported and mixed attributes (numeric and discrete) can be effectively handled. The algorithm evaluates the dependency between class labels and intervals using the χ^2 statistic. Intervals with similar class distributions are merged based on the following formula [23]:

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{k}\frac{\left(A_{ij}-E_{ij}\right)^2}{E_{ij}} \quad (1)$$

k = count of classes
A_{ij} = quantity of patterns in the ith interval, j_{th} class
R_i = total patterns in the i_{th} interval = $\Sigma_{j=1}{}^{k}\, A_{ij}$
C_j = total patterns belonging to the j_{th} class = $\Sigma_{i=1}{}^{2}\, A_{ij}$
N = overall patterns = $\Sigma_{i=1}{}^{2}\, R_i$
E_{ij} = expected value of $A_{ij} = R_i \times \frac{C_j}{N}$

Equation (1) measures the dependency between each feature interval and the class label using the χ^2 statistic. A higher χ^2 value indicates a stronger association between the feature and the target class, meaning that the feature contributes more discriminative information for classification. In this study, the χ^2 criterion was used to rank all candidate features, and the top-ranked 5, 10, 15, and 30 features were retained for subsequent classifier evaluation [23].

2) MRMR

MRMR uses a more sophisticated approach by taking into account two criteria at once, which are maximizing feature relevance to the target class and minimizing redundancy among chosen features [24]. This algorithm is chosen because MRMR is highly effective for high-dimensional datasets as it can identify optimal feature subsets without excessive redundancy. The formula for MRMR is as follows [25]:

$$max X_i \in S\left[I(X_i;c) - \frac{1}{|S|}\sum_{X_j} I(X_j;X_i)\right] \quad (2)$$

Where:

Xi = the *i-th* candidate feature that has not been selected yet (not in the set *S*)
S = The set of features which has been previously selected.
c = the target class/label
Xj = the *j-th* feature contained in *S* (previously chosen)
|S| = the count of features in *S*

Equation (2) shows that MRMR selects features by maximizing their mutual information with the target class while minimizing redundancy among the selected features [25]. In other words, the method favors features that are both highly relevant for distinguishing benign and malignant cases and minimally overlapping in the information they provide. In this study, the MRMR criterion was applied to rank all candidate features, after which the top 5, 10, 15, and 30 features were used as input subsets for the classification models [24].

3) ReliefF

ReliefF is a development of the Relief algorithm designed to handle multi-class data and missing values. This algorithm operates by assessing feature quality, determining how effectively a specific feature can discriminate between instances across different classes [26]. This algorithm is used because ReliefF is capable of detecting interactions between features and can work with various data types without specific distribution assumptions. The core mechanism for estimating feature weights (W) is mathematically formalized as follows [27]:

$$\mathrm{W[A]} = \mathrm{W[A]} - \frac{\mathrm{diff(A, R_i, H)}}{\mathrm{m}} + \frac{\mathrm{diff(A, R_i, M)}}{\mathrm{m}} \quad (3)$$

$$diff(A, X, Y) = \frac{|value(A, X) - value(A, Y)|}{\max(A) - \min(A)} \quad (4)$$

Equation (3) updates the weight of each feature according to its ability to distinguish between neighboring instances from different classes while remaining consistent for instances belonging to the same class. Equation (4) defines the normalized difference function used in the weight update, allowing feature values to be compared on a common scale. Therefore, features that consistently separate malignant from benign samples receive higher weights and higher ranking positions. In this study, the resulting ReliefF weights were used to construct the top 5, 10, 15, and 30 feature subsets for model evaluation [26], [27].

4) Feature Ranking Strategy

Each feature selection algorithm will produce feature rankings based on their respective scores. From these rankings, 4 subsets will be selected: top 5, 10, 15, and 30 features to evaluate their performance. The number of selected features is determined to observe the impact of feature ranking on the computational accuracy of the machine learning model.

D. Machine Learning Classifiers

This section will discuss the working procedures of using DT, SVM, Neural Networks, Gradient Boosted Tree, Gradient Bagged Tree.

1) Decision Trees (DT)

DT is a classification model that computes the output of a function $f(x)$ through a sequential testing process on input x. In this structure, the result of each specific test dictates the subsequent step, proceeding iteratively until the precise value of $f(x)$ is determined [28]. The DT works by selecting the optimal split at each node based on a purity criterion. This model uses the following entropy formula [29]:

$$H(S) = -\Sigma i\ pi\ log2(pi) \quad (5)$$

where pi = the proportion of samples belonging to class *i*. The value of entropy ranges from 0 (pure) to $\log_2$(c), where *c* is the number of classes.

Equation (5) is used to measure the impurity of a node in the decision tree. A lower entropy value indicates that the samples within a node belong predominantly to a single class, whereas a higher entropy value reflects greater class heterogeneity. During training, the decision tree selects splits that maximize entropy reduction, thereby producing more homogeneous child nodes and improving classification performance.

2) Support Vector Machine (SVM)

SVM is a supervised machine learning algorithm used primarily for classification. By analyzing labeled training datasets, the model learns to accurately predict outcomes for new, previously unseen data. SVM operates by identifying a separating hyperplane that establishes the widest possible gap between the closest data points of different categories. This boundary effectively acts as a decision threshold, allowing the model to classify data based on its position relative to the line [30]. Where the equation of the hyperplane is defined as follows [31]:

$$f(x) = w^T x + b \quad (6)$$

where the weight vector is denoted by w, and the bias is represented by b.

Equation (6) defines the separating hyperplane used by SVM to distinguish between benign and malignant classes. The optimal hyperplane is determined by maximizing the margin between support vectors from different classes, thereby improving generalization performance.

3) Neural Networks (NN)

Neural Networks is a computational architecture consisting of interconnected layers of neurons. In this research, the Neural Network model was implemented using the Narrow Neural Network architecture available in MATLAB Classification Learner. This architecture consists of a feedforward neural network with a single hidden layer containing 10 neurons. The

number of input neurons depends on the selected feature subset (5, 10, 15, or 30 features), while the output layer contains two neurons representing benign and malignant classes. The selection of Neural Networks is based on its ability to learn complex non-linear patterns in data and its proven effectiveness in various classification tasks [32].

The learning process is performed using the backpropagation algorithm. The main formula used in weight updating is given as follows [33]:

$$\frac{\partial E}{\partial w} = \frac{\partial E}{\partial o} \times \frac{\partial o}{\partial net} \times \frac{\partial net}{\partial w} \quad (7)$$

where:
E = the error function
w = the weight
o = the output
net = the weighted sum of inputs

Equation (7) represents the weight adjustment process in the backpropagation learning algorithm. During training, the network minimizes the error function by iteratively updating connection weights, enabling the model to learn complex nonlinear relationships within the dataset.

4) Boosted Tree

Boosted Tree is an ensemble method that combines multiple decision trees sequentially for classification tasks. Each new tree is trained to correct prediction errors from the combination of previous trees by minimizing the loss function through gradient descent [34]. The main formula of this model is as follows [35]:

$$F_{m(x)} = F_{\{m-1\}(x)} + \gamma_m h_{m(x)} \quad (8)$$

where:
$F_{\mathrm{m}}(x)$ = the ensemble model at the m-th iteration
γ_{m} is the learning rate
$h_{\mathrm{m}}(x)$ = the weak learner (decision tree) trained to predict the negative gradient of the loss function

Equation (8) describes the ensemble prediction generated during each boosting iteration. The model progressively reduces classification errors by adding weak learners that focus on previously misclassified samples.

5) Bagged Tree

Bagged Tree is an ensemble method that uses bootstrap aggregating techniques to improve classification accuracy. In this ensemble, multiple decision trees are trained on resampled subsets of the dataset, with final predictions determined through majority voting [34]. The main formula of this model is as follows [36]:

$$\varphi_{A(x)} = \left(\frac{1}{B}\right) * \Sigma(b = 1 \; to \; B)\varphi(x, L_b) \quad (9)$$

where:
$\varphi_{A(x)}$ = the aggregated predictor
B = the number of bootstrap samples
φ (x, L_b) = the predictor trained on the b-th bootstrap sample
L_b = the bootstrap sample drawn from the original dataset

Equation (9) represents the aggregation mechanism of bagging, where predictions from multiple bootstrap-trained trees are combined through majority voting to improve stability and reduce overfitting.

E. Hyperparameter Configuration

All classification models were implemented using MATLAB R2024a Classification Learner using the built-in classifier configurations. The Support Vector Machine (SVM) classifier employed a Gaussian kernel function with standardized predictors. The Decision Tree classifier used the default Gini diversity criterion for node splitting. The Neural Network classifier utilized MATLAB's Narrow Neural Network architecture, while Bagged Tree and Boosted Tree classifiers were implemented using the ensemble learning methods available in MATLAB Classification Learner. To ensure a fair comparison, all classifiers were trained and evaluated under identical data partition settings, feature selection methods, and performance evaluation metrics. All model hyperparameters were kept at the default settings provided by MATLAB R2024a Classification Learner unless otherwise specified. For SVM, a Gaussian kernel with standardized predictors was used. The Neural Network model employed MATLAB's Narrow Neural Network architecture with one hidden layer of 10 neurons. The Bagged Tree and Boosted Tree classifiers were implemented using MATLAB's default ensemble settings. Model selection within each training set was based on the average validation performance obtained from 5-fold cross-validation, and the selected model was then evaluated on the corresponding held-out test set.

F. Model Evaluation

Each classification algorithm's performance will be assessed based on the following metrics [36]:

1) Accuracy (A): the ratio of correct predictions to the total number of predictions made. Accuracy formula:

$$\mathrm{A} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \quad (10)$$

2) Precision (PPV): proportion of true positive from total positive predictions. Precision formula:

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \quad (11)$$

3) Specificity (TNR): proportion of true negative from total negative predictions. Specificity formula:

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \quad (12)$$

4) Recall (TPR): proportion of true positive from total actual positive cases. Recall formula:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad (13)$$

5) F1-score (F): harmonic mean of precision and recall. F1-score formula:

$$\mathrm{F} = \frac{2 \times \mathrm{PPV} \times \mathrm{TPR}}{\mathrm{PPV} + \mathrm{TPR}} \quad (14)$$

6) AUC-ROC: total area situated beneath the Receiver Operating Characteristic curve.

Equations (10) - (14) define the evaluation metrics used to compare the performance of all classification models. Accuracy in Equation (10) measures the overall proportion of correct predictions, precision in Equation (11) reflects the reliability of positive predictions, specificity in Equation (12) quantifies the correct identification of benign cases, recall in Equation (13) measures the ability to detect malignant cases, and F1-score in Equation (14) summarizes the balance between precision and recall. These metrics were selected to provide a comprehensive assessment of model performance, particularly because breast cancer classification requires not only high overall accuracy but also strong sensitivity and specificity [36].

IV. Results and Discussions

This part offers an in-depth examination of the results from experiments focused on applying machine learning methods for identifying breast cancer in the fields of healthcare studies and data analytics.. A total of 180 experimental scenarios were systematically designed by varying multiple parameters, including data partition ratios (80:20, 75:25, and 70:30), feature selection techniques (χ^2, MRMR, and ReliefF), the number of chosen features (5, 10, 15, and 30), and classification algorithms (DT, SVM, Neural Networks, Bagged Tree, and Boosted Tree). Model performance was rigorously assessed using several evaluation metrics, namely accuracy, precision (PPV), specificity (TNR), recall (TPR), and F1-score. All experiments were implemented using MATLAB R2024a to identify the most optimal combination of feature selection strategies and classification models for breast cancer diagnosis.

A. Algorithm Performance Comparison

Table III presents the best performance achieved by each algorithm across all experimental configurations. SVM achieved the highest accuracy of 97.81% using MRMR feature selection with 15 features and 80:20 data split, followed by Neural Networks with 97.42% accuracy.

TABLE III. Performance Comparison of Machine Learning Algorithms

Algorithm	FS	No. Features	Train: Test	Acc	PPV	TNR	F1-score	Recall
SVM	MRMR	15	80:20	97.81	99.38	99.65	96.99	94.71
Neural Network	ReliefF	30	75:25	97.42	98.05	98.88	96.49	94.97
Bagged Tree	MRMR	30	80:20	95.83	96.32	97.90	94.29	92.35
Decision Tree	χ^2	5	80:20	94.08	91.81	95.10	92.08	92.35
Boosted Tree	MRMR	5	75:25	75.18	93.44	98.51	51.82	35.85

SVM consistently demonstrated superior performance across different configurations. The algorithm achieved high precision (99.38%) and specificity (99.65%), indicating excellent ability to correctly identify both malignant and benign cases. Neural Networks showed competitive performance with balanced metrics across all evaluation criteria.

B. Performance Comparison Based on Feature Selection Methods

Table IV compares the effectiveness of different feature selection methods under the 80:20 data split in achieving optimal performance of each algorithm.

TABLE IV. Accuracy Comparison of Feature Selection Methods Across Algorithm.

Algorithm	χ^2(%)	MRMR(%)	ReliefF(%)	Best Feature Selection
SVM	97.81(30)	**97.81(15)**	97.81(30)	MRMR with 15 parameters
Neural Networks	**96.49(30)**	96.05(30)	96.27(15)	χ^2 with 30 parameters
Bagged Tree	94.52(30)	**95.83(30)**	95.39(5)	MRMR with 30 parameters
DT	**94.08(5)**	93.64(30)	93.42(5)	χ^2 with 5 parameters
Boosted	69.08(15)	69.30(5)	**69.52(30)**	ReliefF with 30 parameters

MRMR demonstrated superior efficiency by achieving peak accuracy with only 15 features, representing a 50% reduction compared to the full feature set. All three feature selection methods achieved the same peak accuracy (97.81%) when combined with SVM, indicating the robustness of the algorithm.

Note that Table IV is restricted to the 80:20 split; the overall best Neural Networks result (97.42% with ReliefF) in Table III occurs at the 75:25 split.

C. Algorithm Performance Analysis

1) Support Vector Machine (SVM)

SVM demonstrated superior and consistent performance compared to other algorithms. SVM's highest accuracy reached 97.81% with various feature selection configurations. Since SVM separates the two classes by searching for the widest margin between them, it worked well on the standardized features of this dataset. Its precision (99.38%) and specificity (99.65%) were high because very few benign cases were predicted as malignant. The accuracy of SVM also changed little across the three train–test ratios, showing that it was not very sensitive to the amount of training data.

2) Neural Networks (NN)

Neural Networks showed very good performance with the highest accuracy of 97.42% in the configuration of top 30 features with ReliefF. Neural Networks is capable of capturing complex patterns in data, but requires more careful parameter tuning compared to SVM.

3) Decision Tree (DT)

DT provided good performance with average accuracy above 90%. Although it did not achieve accuracy as high as SVM or Neural Networks, DT still provided acceptable results with the highest accuracy of 94.08% on the top 5 features with χ^2

4) Bagged Tree

Bagged Tree showed improved performance compared to single DT with the best accuracy of 95.83% in several configurations. This ensemble method is effective in reducing overfitting and improving model generalization.

5) Boosted Tree

Boosted Tree showed suboptimal performance with very low accuracy (around 62-75%). The confusion matrices explain why. In most configurations the model predicted almost everything as benign, producing very few true positives and a recall near zero. An accuracy of 62% here mainly reflects the share of benign cases in the data rather than any real ability to detect malignant tumours. Even the best run, 75.18% with MRMR and five features, reached a recall of only 35.85%, meaning many malignant cases were missed. Such a result is not acceptable for screening. Since the other four classifiers handled the same data well, this failure is specific to the boosting configuration rather than a property of the dataset. The most likely cause is the default ensemble setting in MATLAB, which was not suited to this case, and not a weakness of boosting as a method, so the result should not be generalized. Tuning the number of weak learners and the learning rate would be expected to improve it.

D. Feature Selection Performance Analysis

1) Chi-square Feature Selection

Evaluation results using the Chi-square method showed that SVM had the best performance with accuracy up to 97.81% on the top 30 features (80:20 data split). In this configuration, SVM achieved precision of 98.78%, specificity of 99.30%, F1-score of 97.01%, and recall of 95.29%. Neural Networks also showed good performance with 96.49% accuracy on the top 30 features, while DT achieved maximum accuracy of 94.08% on the top 5 features.

2) MRMR Feature Selection

The MRMR approach delivered excellent performance, with the SVM model reaching its highest accuracy of 97.81% when using the top 15 features under the 80:20 data split. These results demonstrate MRMR's efficiency in selecting relevant features with minimum redundancy. Under this setup, the SVM model recorded strong metric performance, attaining a precision of 99.38%, a specificity of 99.65%, an F1-score of 96.99%, and a recall rate of 94.71%. Neural Networks with MRMR achieved 96.96% accuracy at the 75:25 data split with the top 30 features.

3) ReliefF Feature Selection

ReliefF showed competitive performance with SVM achieving 97.81% accuracy on the top 30 features (80:20 data split). Neural Networks with ReliefF demonstrated very good performance with 97.42% accuracy at the 75:25 data split with the top 30 features, achieving precision of 98.05%, specificity of 98.88%, F1-score of 96.49%, and recall of 94.97%.

V. Conclusion and Future Works

Recent advances in computational intelligence have highlighted the significant role of machine learning cancer diagnosis. In this study, the WDBC was utilized to conduct a comprehensive evaluation of several classification algorithms, namely SVM, DT, Neural Networks, Boosted Tree, and Bagged Tree. These classifiers were systematically integrated with three feature selection approaches, MRMR, χ^2, and ReliefF, across multiple feature subset configurations (5, 10, 15, and 30 features).

The empirical findings demonstrate that SVM consistently showed better performance compared to the other models for every evaluation measure, achieving a maximum accuracy of 97.81% when 15 and 30 features were employed. Notably, this peak accuracy was reached by MRMR using only 15 of the 30 features, halving the dimensionality without any loss in accuracy, which is useful for building lighter computer-aided diagnosis systems. In contrast, The Bagged Tree model also showed strong performance, reaching its highest accuracy of 95.83% with MRMR using 30 selected features. In a reduced-feature setting, the model remained competitive, achieving 95.39% with ReliefF using only five features. These outcomes underscore the critical importance of aligning appropriate feature selection strategies with suitable classification algorithms to optimize predictive performance in breast cancer diagnosis. Within the scope of this investigation, SVM, Neural Networks, and Bagged Tree models were identified as the most effective configurations.

The results further provide methodological implications for subsequent research. While a conventional Neural Network architecture was implemented, more sophisticated deep learning models, such as Convolutional Neural Networks (CNNs), may yield enhanced predictive capability, particularly when complemented by data augmentation techniques. Despite the high performance achieved by SVM, further refinement remains feasible. Enhanced model performance may be realized through more rigorous hyperparameter optimization, including systematic tuning of kernel functions (e.g., linear, radial basis function, or polynomial), the regularization parameter

(C), and the gamma parameter. Additionally, the incorporation of dimensionality reduction techniques such as PCA, as well as imbalance-handling methods like SMOTE, may further strengthen generalization performance and model robustness.

ACKNOWLEDGMENT

The authors gratefully acknowledge the Department of Medical Technology, Institut Teknologi Sepuluh Nopember, for providing the facilities and institutional support that contributed to the completion of this research.

REFERENCES

[1] [1 Y. Zhang, Y. Ji, S. Liu, J. Li, J. Wu, Q. Jin, X. Liu, H. Duan, Z. Feng, Y. Liu, Y. Zhang, Z. Lyu, F. Song, F. Song, L. Yang, H. Liu, and Y. Huang, "Global burden of female breast cancer: new estimates in 2022, temporal trend and future projections up to 2050 based on the latest release from GLOBOCAN," J. Natl. Cancer Cent., vol. 5, no. 3, pp. 287–296, 2025, doi: https://doi.org/10.1016/j.jncc.2025.02.002.

[2] A. C. Society, "Breast Cancer Facts & Figures 2023-2024," American Cancer Society, Atlanta, 2023. [Online]. Available: https://www.cancer.org/research/cancer-facts-statistics/breast-cancer-facts-figures.html

[3] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, "Cancer statistics, 2023.," CA. Cancer J. Clin., vol. 73, no. 1, pp. 17–48, Jan. 2023, doi: 10.3322/caac.21763.

[4] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries.," CA. Cancer J. Clin., vol. 71, no. 3, pp. 209–249, May 2021, doi: 10.3322/caac.21660.

[5] W. H. Organization, "Breast cancer," World Health Organization, Mar. 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/breast-cancer

[6] P. H. Tan, I. Ellis, K. Allison, E. Brogi, S. B. Fox, S. Lakhani, A. J. Lazar, E. A. Morris, A. Sahin, R. Salgado, A. Sapino, H. Sasano, S. Schnitt, C. Sotiriou, P. van Diest, V. A. White, D. Lokuhetty, and I. A. Cree, "The 2019 World Health Organization classification of tumours of the breast," Histopathology, vol. 77, no. 2, pp. 181–185, 2020, doi: 10.1111/his.14091.

[7] W.-C. Shia, L.-S. Lin, and D.-R. Chen, "Classification of malignant tumours in breast ultrasound using unsupervised machine learning approaches," Sci. Rep., vol. 11, no. 1, p. 1418, 2021, doi: 10.1038/s41598-021-81008-x.

[8] M. Arnold, E. Morgan, H. Rumgay, A. Mafra, D. Singh, M. Laversanne, J. Vignat, J. R. Gralow, F. Cardoso, S. Siesling, and I. Soerjomataram, "Current and future burden of breast cancer: Global statistics for 2020 and 2040.," Breast, vol. 66, pp. 15–23, Dec. 2022, doi: 10.1016/j.breast.2022.08.010.

[9] J. T. Henderson, E. M. Webber, M. S. Weyrich, M. Miller, and J. Melnikow, "Screening for Breast Cancer: Evidence Report and Systematic Review for the US Preventive Services Task Force.," JAMA, vol. 331, no. 22, pp. 1931–1946, Jun. 2024, doi: 10.1001/jama.2023.25844.

[10] F. Cucchiara, R. Scarpitta, S. Crucitta, C. Scatena, R. Arici, A. G. Naccarato, S. Fogli, R. Danesi, and M. Del Re, "Diagnosis and treatment monitoring in breast cancer: how liquid biopsy can support patient management.," Pharmacogenomics, vol. 23, no. 2, pp. 119–134, Jan. 2022, doi: 10.2217/pgs-2021-0099.

[11] K. Loizidou, R. Elia, and C. Pitris, "Computer-aided breast cancer detection and classification in mammography: A comprehensive review," Comput. Biol. Med., vol. 153, p. 106554, 2023, doi: 10.1016/j.compbiomed.2023.106554.

[12] H. Asri, H. Mousannif, H. Al Moatassime, and T. Noel, "Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis," Procedia Comput. Sci., vol. 83, pp. 1064–1069, 2016, doi: https://doi.org/10.1016/j.procs.2016.04.224.

[13] 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. Updat., vol. 3, p. 100098, 2023, doi: https://doi.org/10.1016/j.cmpbup.2023.100098.

[14] S. Ara, A. Das, and A. Dey, "Malignant and Benign Breast Cancer Classification using Machine Learning Algorithms," in 2021 International Conference on Artificial Intelligence (ICAI), 2021, pp. 97–101. doi: 10.1109/ICAI52203.2021.9445249.

[15] A. Khalid, A. Mehmood, A. Alabrah, B. F. Alkhamees, F. Amin, H. AlSalman, and G. S. Choi, "Breast Cancer Detection and Prevention Using Machine Learning.," Diagnostics (Basel, Switzerland), vol. 13, no. 19, Oct. 2023, doi: 10.3390/diagnostics13193113.

[16] S. Aamir, A. Rahim, Z. Aamir, S. F. Abbasi, M. S. Khan, M. Alhaisoni, M. A. Khan, K. Khan, and J. Ahmad, "Predicting Breast Cancer Leveraging Supervised Machine Learning Techniques," Computational and Mathematical Methods in Medicine, vol. 2022, Article ID 5869529, 2022, doi: 10.1155/2022/5869529.

[17] E. Strelcenia and S. Prakoonwit, "Effective Feature Engineering and Classification of Breast Cancer Diagnosis: A Comparative Study," BioMedInformatics, vol. 3, no. 3, pp. 616-631, 2023, doi: 10.3390/biomedinformatics3030042.

[18] M. A. Naji, S. El Filali, K. Aarika, E. L. H. Benlahmar, R. A. Abdelouhahid, and O. Debauche, "Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis," Procedia Comput. Sci., vol. 191, pp. 487-492, 2021, doi: https://doi.org/10.1016/j.procs.2021.07.062.

[18] Al-Imran, S. Akter, M. A. S. Mozumder, R. J. Bhuiyan, T. Rahman, M. J. Ahmmed, M. N. H. Mir, M. A. Hasan, A. C. Das, and M. E. Hossen, "Evaluating Machine Learning Algorithms for Breast Cancer Detection: A Study on Accuracy and Predictive Performance," Am. J. Eng. Technol., vol. 6, no. 09, pp. 22-33, 2024, doi: 10.37547/tajet/Volume06Issue09-04.

[19] A. Rasool, C. Bunterngchit, L. Tiejian, M. R. Islam, Q. Qu, and Q. Jiang, "Improved Machine Learning-Based Predictive Models for Breast Cancer Diagnosis," International Journal of Environmental Research and Public Health, vol. 19, no. 6, art. 3211, 2022, doi: 10.3390/ijerph19063211.

[20] O. O. Okundalaye and N. Özdemir, "Early breast cancer prediction using optimized machine learning and tumor-immune modeling," Journal of Computational and Applied Mathematics, vol. 473, art. 116875, 2026, doi: 10.1016/j.cam.2025.116875.

[21] T. Arifin, I. W. P. Agung, E. Junianto, R. Rachman, I. R. Wibowo, and D. D. Agustin, "Breast cancer identification using machine learning and hyperparameter optimization," Indones. J. Electr. Eng. Comput. Sci., vol. 36, no. 3, pp. 1620–1630, 2024, doi: 10.11591/ijeecs.v36.i3.pp1620-1630.

[22] R. Setiono and H. Liu, "Chi2: Feature Selection and Discretization of Numeric Attributes," in 2012 IEEE 24th International Conference on Tools with Artificial Intelligence, 1995, p. 388. doi: 10.1109/TAI.1995.479783.

[23] I. Jo, S. Lee, and S. Oh, "Improved Measures of Redundancy and Relevance for mRMR Feature Selection," 2019. doi: 10.3390/computers8020042.

[24] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 8, pp. 1226–1238, 2005, doi: 10.1109/TPAMI.2005.159.

[25] O. Reyes, C. Morell, and S. Ventura, "Scalable extensions of the ReliefF algorithm for weighting and selecting features on the multi-label learning context," Neurocomputing, vol. 161, pp. 168–182, 2015, doi: https://doi.org/10.1016/j.neucom.2015.02.045.

[26] M. Robnik-Šikonja and I. Kononenko, "Theoretical and Empirical Analysis of ReliefF and RReliefF," Mach. Learn., vol. 53, no. 1, pp. 23–69, 2003, doi: 10.1023/A:1025667309714.

[27] H. Blockeel, L. Devos, B. Frénay, G. Nanfack, and S. Nijssen, "Decision trees: from efficient prediction to responsible AI," Front. Artif. Intell., vol. Volume 6-, 2023, doi: 10.3389/frai.2023.1124553.

[28] C. E. Shannon, "A Mathematical Theory of Communication," Bell Syst. Tech. J., vol. XXVII, no. 3, pp. 379–423, 1948.

[29] I. Bomze, F. D'Onofrio, L. Palagi, and B. Peng, "Feature selection in linear support vector machines via a hard cardinality constraint: A scalable conic decomposition approach," Eur. J. Oper. Res., 2025, doi: https://doi.org/10.1016/j.ejor.2025.03.017.

[30] C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995, doi: 10.1007/BF00994018.

[31] E. Bollt, "How neural networks work: Unraveling the mystery of randomized neural networks for functions and chaotic dynamical systems," Chaos An Interdiscip. J. Nonlinear Sci., vol. 34, no. 12, p. 123164, Dec. 2024, doi: 10.1063/5.0206840.

[32] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," Nature, vol. 323, no. 6088, pp. 533–536, 1986, doi: 10.1038/323533a0.

[33] A. Plaia, S. Buscemi, J. Fürnkranz, and E. L. Mencía, "Comparing Boosting and Bagging for Decision Trees of Rankings," J. Classif., vol. 39, no. 1, pp. 78–99, 2022, doi: 10.1007/s00357-021-09397-2.

[34] J. H. Friedman, "Greedy function approximation: A gradient boosting machine.," Ann. Stat., vol. 29, no. 5, pp. 1189–1232, Oct. 2001, doi: 10.1214/aos/1013203451.

[35] L. Breiman, "Bagging predictors," Mach. Learn., vol. 24, no. 2, pp. 123–140, 1996, doi: 10.1007/BF00058655.

[36] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," Information Processing & Management, vol. 45, no. 4, pp. 427–437, 2009, doi: 10.1016/j.ipm.2009.03.002.

[37] A. Elaraby, A. Saad, H. Elmannai, M. Alabdulhafith, M. Hadjouni, and M. Hamdi, "An approach for classification of breast cancer using lightweight deep convolution neural network," Heliyon, vol. 10, no. 20, p. e38524, Oct. 2024, doi: 10.1016/j.heliyon.2024.e38524.