

A Hybrid Weighted Soft-Voting Ensemble with Integrated Preprocessing for Breast Cancer Classification on the WDBC Dataset

Mohammad Khalaf Rahim Al-Juaifari, Israa Mohammed Rahi Alabudy,
Ammar Rasoul Mohammad Alsaachi
College of Medicine, University of Kufa, Najaf, Iraq

ARTICLE INFORMATION

Article History:

Received 27 August 2025
Revised 21 December 2025
Accepted 10 June 2026

Keywords:

Breast Cancer Classification;
Machine Learning;
Ensemble Learning;
Hybrid Model;
XGBoost;
Precision Medicine;
Diagnostic Pipeline

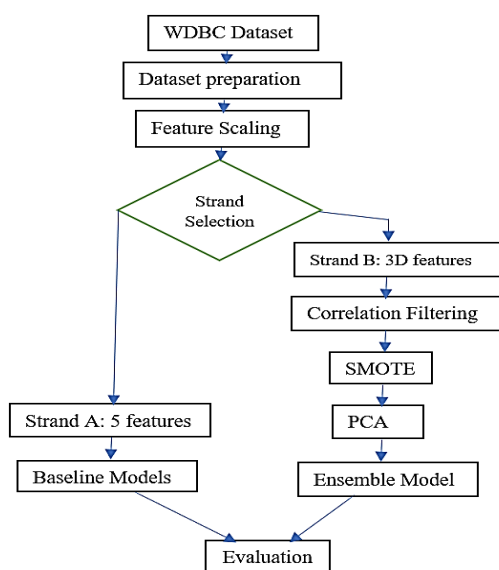
Corresponding Author:

Israa Mohammed Rahi Alabudy,
College of Medicine, University of
Kufa, Najaf, Iraq.
Email:
israam.alabudy@uokufa.edu.iq

This work is open access under a
[Creative Commons Attribution-Share
Alike 4.0](https://creativecommons.org/licenses/by-sa/4.0/)



ABSTRACT



Breast cancer continues to be one of the major causes of deaths due to cancer amongst women around the globe, requiring an effective method of diagnosis. In this research, a machine learning model pipeline that uses a hybrid weighted soft-voting based ensemble method on breast tumor classification for binary data is presented using the Wisconsin Diagnostic Breast Cancer Dataset (WDBC). Data preprocessing includes balancing the classes using Synthetic Minority Over-sampling Technique (SMOTE), removing the highly correlated attributes, and reducing dimensionality through Principal Component Analysis (PCA). Using stratified 10-fold cross-validation, an ensemble yielded 99.3% accuracy and 100% recall for the malignant class on PCA-transformed data, which was better than any individual classifier. Logistic Regression had good performance too, with 98.83% accuracy and 99.81% ROC AUC, which shows that our data can be nearly linearly separable. The feature importance analysis showed that “worst concave points” and “mean radius” were the most important features, and this makes sense from a medical perspective. Overall, this work presents an effective methodology for diagnosing breast cancer.

Document Citation:

M. K. R. Al-Juaifari, I. M. R. Alabudy, and A. R. M. Alsaachi, “A Hybrid Weighted Soft-Voting Ensemble with Integrated Preprocessing for Breast Cancer Classification on the WDBC Dataset,” *Buletin Ilmiah Sarjana Teknik Elektro*, vol. 8, no. 3, pp. 781-791, 2026, DOI: 10.12928/biste.v8i3.14587.

1. INTRODUCTION

As the most common cancer diagnosed and the leading cause of cancer-related death for women globally, breast cancer is a major global health concern [WHO source]. The continuous search for trustworthy diagnostic techniques in oncology and public health is motivated by the fact that early and accurate identification is a crucial factor in determining prognosis and survival rates. Despite their great value, conventional methods like mammography and histological examination have drawbacks such as lower sensitivity in identifying early-stage cancers, subjective interpretation, and inter-observer variability. These limitations may cause patients to undergo needless invasive procedures or dangerous delays in diagnosis. As a result, there is an urgent clinical need for data-driven, intelligent decision-support technologies to enhance medical knowledge and enhance patient outcomes.

Medical diagnostics is changing as a result of advances in artificial intelligence (AI). AI systems are able to examine large amounts of complicated medical data and find trends and insights that would be missed by human observers. Most importantly, supervised learning approaches such as RF, SVM, as well as advanced gradient boosting methods such as XGBoost, seem very promising when used to automate the differentiation process for benign vs. malignant breast tumors [1]-[3]. The main advantage of such approaches is that they are fast, unbiased, and easily reproducible since they leverage previously accumulated data.

Despite all that, there are specific challenges associated with using ML in a healthcare context that might adversely affect the performance and utility of the model. Imbalanced data is an issue where the training dataset contains more examples of healthy patients than sick patients; therefore, the model becomes outperformed in identifying normal cases while overlooking disease since it did not encounter enough examples of it during the training process, as well as feature redundancy and multicollinearity between diagnostic features that will add noise to the data. Other common problems that should be addressed via thorough testing frameworks include generalization of a model from a single dataset and overfitting. While numerous works examine the performance of various classification algorithms separately [3], very few works discuss the entire pipeline process required for building a robust classifier that could be applied in a clinical environment.

The development of the research is marked by efforts such as the advanced ensemble model proposed by Al Reshan *et al.* [4] that combines RFE with bagging, boosting, stacking, and voting. Other research efforts in this area include [5], where the treatment of ordered, categorical data is explored using approaches such as ordinal encoding.

In response to the above-mentioned research gap, this research paper offers a comprehensive and robust machine learning framework for breast cancer diagnosis. Going beyond mere comparison of different models, our study introduces an advanced framework that combines effective data pre-processing techniques with an innovative approach to ensemble learning. The key objectives of the research include increased diagnostic accuracy, resistance to data irregularities, and a scalable model for application in other biomedical tasks.

Specifically, this project makes four primary contributions:

1. Unified Data Preprocessing done by develop a systematic data preprocessing process, taking into consideration various data quality problems that are addressed using SMOTE, correlation filtering and PCA methods.
2. The proposed Hybrid Ensemble develop and test a novel approach to constructing a hybrid ensemble, based on the soft-voting weighted combination method, and utilizing five different algorithms, namely Random Forest, Gradient Boosting, XGBoost, SVC, and logistic regression, which allow one to better understand stable good performance.
3. Extensive Empirical Comparison proposed ensemble was carefully evaluated using a large number of measures to compare its performance with other established baseline classifiers. Then performed rigorous validation techniques to make sure our results were statistically reliable.
4. Clinical Relevance of Features and Model in order to improve credibility and applicability in clinical settings, we conducted a comprehensive feature selection process and evaluated whether or not features used for classification were consistent with known pathological data.

1.1. Data Source and Description

The study utilises the Wisconsin Diagnostic Breast Cancer (WDBC) dataset, which is one of the benchmark datasets in machine-learning research for cancer studies [1]. The dataset consists of 569 FNA image records of tumours found in the breast and can be easily acquired from popular data science packages like Scikit-learn. In this instance, each record is represented by a feature vector based on the numerical values that represent important attributes, such as size, perimeter, and texture, among others, of the tumour nuclei found in the digital images. The problem to solve here is to classify a record as either benign (negative) or malignant (positive). Class imbalance can be observed in this dataset; out of 569 records, 37.2 % of them are malignant,

while 62.7 % of them are benign cases. In order to prevent the biasness towards one class due to imbalance, important preprocessing techniques need to be performed.

1.2. Feature Sets and Experimental Design

Feature configurations of two types from WDBC dataset were used within a double strategy framework of experimentation. With this kind of approach, it became possible to analyze the impact that the quantity of available data has on the trade-off between model simplicity, readability and precision.

1.2.1. Feature-Engineered Subset (First Strategy)

This approach considers a very concise but extremely effective set of predictors. Five predictors including mean radius, mean texture, mean perimeter, mean area and mean smoothness have been selected owing to their known clinical significance and vital role in tumor classification. This well-defined selection process maintains the entire dataset of 569 instances but narrows down the number of predictors to just five most important predictors to fully examine the model performance under the condition of limited availability of information.

1.2.2. Comprehensive Feature Set (Second Strategy)

The second approach makes use of all information available within the WDBC data set through the employment of all thirty original predictor variables. For each cell parameter (for example, texture or radius), predictor variables can be categorized into three different statistical groups as follows: the Mean describes the average measure; the Standard Error (SE) measures dispersion or variability; the “worst” group refers to the average of the most three extreme measures, normally depicting abnormalities.

Examples include mean radius, SE of texture, and worst perimeter. A total of 569×30 matrix is thus generated, which provides a good foundation for assessing the performance of the model with full information discrimination power at the same time posing difficulties associated with high dimensionality and multicollinearity. The reason why a dual approach is used is to perform an all-encompassing examination of algorithmic performance in different situations to see whether a complex model will yield additional information or whether a simple and clinically significant selection of predictors is enough to yield accurate results. The defining properties of these two data sets are shown in [Table 1](#).

Table 1. Summary of Dataset Configurations

Configuration	Instances	Features	Class Distribution (Benign:Malignant)	Primary Objective
Feature-Engineered Subset	569	5	357:212	Evaluate efficiency and performance with minimal, clinically-relevant features. Maximize predictive accuracy using the full spectrum of available data.
Comprehensive Feature Set	569	30	357:212	

1.3. Related Work

The application of ML in diagnosing breast cancer has been one of the most vibrant fields of research for several decades. This is because of the pressing need to detect diseases in their early stages and also because of the existence of publicly available benchmark datasets, e.g., the Wisconsin Breast Cancer Dataset (WDBC). Current studies conducted in this realm fall into three main streams: those that study the behavior of classifiers individually, those concerned with feature generation, and those centered around ensemble learning methods.

1.3.1. Individual Classifiers and Feature Engineering

An important part of the literature in this field has been dedicated to the evaluation and benchmarking of the performance of stand-alone machine learning classifiers. In their research, Strelcenia and Prakoonwit [6] performed a comparative analysis of six classical algorithms on the WBCD, using substantial amounts of feature engineering. By obtaining a classification accuracy of 98.64%, the method introduced in [6] clearly demonstrates the importance of choosing an appropriate set of features. The use of hybrid architectures combining both deep and traditional machine learning approaches was presented in the work of [7].

In particular, the authors used 1D-CNN feature extraction for subsequent classification with some classical algorithms, thus obtaining 98.24% test accuracy by using XGBoost along with 1D-CNN features. Analyzing eight different algorithms on the same dataset, Hossin *et al.* [3] demonstrated the best performance of Logistic Regression, reaching 99.12%. This can be corroborated by the work done by Khater *et al.* [8], who used Random Forest in an XAI setting and showed that this model was not only highly accurate but also had clear decision-making processes essential for clinical application. In another study [6], researchers integrated patient details along with mammogram images through a CNN model with 23,467 patient data tested.

1.3.2. Advanced and Ensemble Learning Techniques

In order to circumvent some of the shortcomings associated with single classifiers, various models have become more complicated with time. Albadr *et al.* [2] proposed a Fast-Learning Network (FLN) which was able to achieve an accuracy level of 98.37% when applied to the WBCD dataset while also showing high flexibility and low susceptibility to overfitting. Another example is presented in Gurcan [9], where a stacking algorithm created an ensemble using the LightGBM, CatBoost, and CNN algorithms, resulting in a 97.72% accuracy rate.

Apart from traditional machine learning techniques, modern studies have turned to adjacent technologies to boost classifier performance. One such study by Kavitha *et al.* [10] involved modifying the YOLOv3 computer vision framework, which achieved 97.21% accuracy in breast cancer diagnosis while exemplifying the possibilities of transfer learning in the field. Another direction involves applying techniques from other areas to solve data access problems. This way, Shukla *et al.* [11] have used Differential Privacy in combination with Federated Learning to train models using distributed data without the exchange of actual records, achieving a 96.1% accuracy in breast-cancer detection.

1.3.3. Identification of the Research Gap

Nevertheless, this advancement is still accompanied by some shortcomings present in previous studies. Namely, either researchers analyse individual classifiers without creating an end-to-end pre-processing technique, or, conversely, they focus on advanced ensemble design while ignoring crucial data aspects like redundant features and imbalanced classes.

In light of this, there is an evident research gap regarding designing an end-to-end data processing methodology which integrates both comprehensive data preparation through advanced resampling techniques and dimensionality reduction methods as well as the construction of an ensemble from various base learning methods. It is this research gap that our study addresses by proposing a holistic approach that involves robust pre-processing along with a weighted voting-based heterogeneous ensemble in conjunction with two types of feature selection.

While RF, SVM, GB, and XGBOOST have all demonstrated superior performance for WDBC in their own way, each suffers from inherent weaknesses when deployed independently. Ensemble methods like RF, GB, and XGB produce high accuracy but tend to be unstable with small imbalanced datasets; logistic regression gives good interpretability but tends to underfit complex decision boundaries; while the maximum-margin classifier, SVC, is unstable with different feature scales and imbalances.

Existing hybrid architectures [4],[9]-[12] utilize various combinations of these machine learning methods but fail to employ a coherent preprocessing pipeline for handling multicollinearity, class imbalance, and dimensionality reduction. What makes our proposal novel is that we (i) use empirical weights of complementary base learners for better prediction stability and (ii) introduce an end-to-end conditioning module for biomedical data (SMOTE + correlation filter + PCA).

The rest of the paper is structured as follows. In Section 2, we give details about our proposed approach and the used dataset. Experiments with the dataset and discussion about the results and comparison with existing literature are included in Section 3. Finally, we draw conclusions in Section 4.

2. MATERIALS AND METHODS

This paper aimed at developing an effective machine learning pipeline for the classification of breast cancer tumors. As can be seen in Figure 1, the general methodology followed by the paper has been structured into various stages, with two strands running in parallel to investigate the effect of feature selection on the process. The steps undertaken were: 1. WDBC Dataset→Data Preprocessing→Two Strands (A & B), 2. Strand A: 5 features→Baseline models→Evaluation and, 3.Strand B: 30 features→Advanced Data Preprocessing→Ensemble→Evaluation.

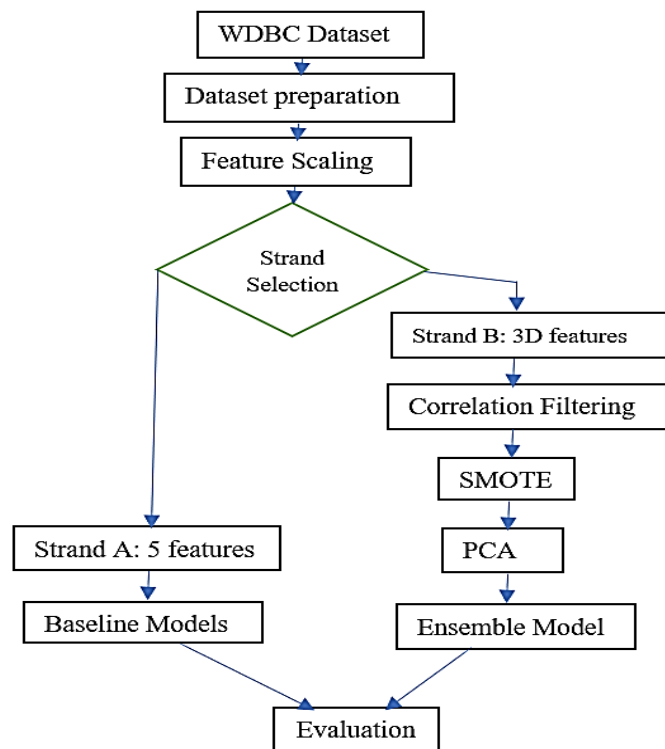


Figure 1. Methodology Workflow Diagram

2.1. Dataset Description

The Wisconsin Diagnostic Breast Cancer (WDBC) database was downloaded from the UCI Machine Learning Repository. This database consists of 569 fine-needle aspirate (FNA) biopsies, where each biopsy comprises 30 continuous attributes extracted from digitized images of breast mass nuclei. Classification for this database is binary with malignant cancer (212 cases, 37.2%) or benign cancer (357 cases, 62.8%).

2.2. Experimental Design: Dual-Strand Approach

Two experiment strands were carried out to conduct an all-round assessment of the model performances based on varying levels of information availability: Strand A – Feature engineered subset, which used only five clinically important features (average radius, average texture, average perimeter, average area, average smoothness), to represent a case of limited information availability. Strand B–Comprehensive set of 30 features.

2.3. Data Pre-processing Pipeline

Preprocessing stages were implemented by scikit-learn (version 1.3.0) with firm parting between training and test data to avoid information leakage.

2.3.1. Common Preprocessing Steps (Both Strands):

There are three main steps: a) **Label Encoding** is categorical labels ('M' for malignant, 'B' for benign) were mapped to numerical values (1 and 0, respectively), b) **Train-Test Split** is data were divided using stratified sampling (70% training, 30% testing, random_state=42) to preserve class distribution, and c) **Feature Scaling** were standardized to zero mean and unit variance using StandardScaler, fit exclusively on the training set.

2.3.2. Strand-Specific Preprocessing:

For Strand A (five features), Only the five chosen features were used after scaling, and for Strand B (30 features): 1. **Correlation Filtering:** The features having pairwise correlation greater than 0.95 were discarded in order to avoid multicollinearity. 2. **Class Balancing:** SMOTE was employed only in the training set in order

to solve the problem of imbalanced classes, and 3. Dimensionality Reduction: PCA, which retains 95% of variance was performed.

2.4. Model Development and Tuning

2.4.1. Baseline Models

Three advanced algorithms were built and optimized through 5-fold cross validation as follows: 1) Logistic regression (LR) with L2 regularization, where C was optimized using grid search, 2) Random Forest (RF), where number of estimators = 100, and max depth optimized using grid search, and 3) XGBoost (XGB).

2.4.2. Proposed Hybrid Weighted Soft-Voting Ensemble

The optimization process involved maximizing ROC-AUC, with the final models being scored using the test dataset. The proposed approach used a new type of heterogeneous ensemble learning algorithm that uses five base estimators, namely Random Forest Classifier, Gradient Boosting Classifier, XGBoost Classifier, Support Vector Classifier (probability=True), and Logistic Regression. The method makes use of weighted soft voting whereby the predictions are made by considering the weighted average of class probabilities obtained from each of the base learners. The weights assigned to the base models include [3, 2, 3, 1, 1].

2.4.3. Hyperparameter Optimization

Models experienced hyperparameter tuning via GridSearchCV with 5-fold stratified cross-validation. The parameter grid for XGBoost comprised: n_estimators: [100, 200, 300], max_depth: [3, 5, 7], learning_rate: [0.01, 0.1], subsample: [0.8, 1.0], and colsample_bytree: [0.8, 1.0].

2.5. Evaluation Framework

The performance of models was measured based on different measures on an independent test set. 10-Fold cross-validation along with five repetitions added to the robustness of the evaluation process. A) Accuracy: General accuracy of classification, B) Precision, Recall, F1 Score: Performance in classes separately, C) ROC-AUC: Threshold-independent discriminatory power, D) Log-Loss: Calibration of probabilistic predictions, and E) Confidence Intervals: 95% CIs by bootstrapping ($n = 1000$).

3. RESULTS AND DISCUSSION

This segment is dedicated to discussion of experimental results for the proposed hybrid ensemble system. This segment will be divided into five parts for clarity purposes: Performance on the engineered feature subset, performance on the full feature set, model comparison, feature importance and feature correlation analysis, and comparison with existing work.

3.1. Performance on Feature-Engineered Subset (5 Features)

To evaluate model effectiveness under controlled information, first employed only five clinically important features. The hybrid ensemble attained strong performance on both original and PCA-transformed features, as summarized in Table 2, Values represent mean \pm 95% confidence interval from 10-fold cross-validation. The use of PCA transformation greatly increased the effectiveness, especially through obtaining perfect recall (100 percent) for malignant tumors which was essential in the application of PCA since missing malignancies could be fatal.

Table 2. Performance Comparison: Original vs. PCA-Transformed Features (5-Feature Subset)

Metric	Original Features	PCA-Transformed Features
Accuracy	95.8% ($\pm 1.2\%$)	99.3% ($\pm 0.8\%$)
Balanced Accuracy	95.8% ($\pm 1.3\%$)	99.3% ($\pm 0.9\%$)
Precision (Malignant)	94.5% ($\pm 1.8\%$)	98.6% ($\pm 1.1\%$)
Recall (Malignant)	97.1% ($\pm 1.5\%$)	100% ($\pm 0.0\%$)
F1-Score	95.8% ($\pm 1.2\%$)	99.3% ($\pm 0.8\%$)
ROC AUC	0.9965 (± 0.002)	0.9961 (± 0.002)
Log Loss	0.0871 (± 0.012)	0.0813 (± 0.011)

3.2. Performance on Comprehensive Feature Set (30 Features)

The features of 30 original with full pre-processing represents by (correlation filtering, SMOTE, PCA), by evaluating both baseline models and the proposed ensemble method in this paper, Table 3 presents the comparative performance with baseline models, values represent mean \pm 95% confidence interval from 5×10 -fold cross-validation. It is also worth noting that Logistic Regression had the best Accuracy and ROC AUC

scores at 98.83% and 99.81%, respectively. This result implies that the WDBC data set is almost linearly separable. The implication of this result is that linear models can work extremely well when conditioned appropriately.

Table 3. Comparative Performance of Baseline Models and Proposed Ensemble (30 Features)

Model	Accuracy	ROC AUC	Precision	Recall	F1-Score	Key Characteristics
Logistic Regression	98.83% ($\pm 0.9\%$)	99.81% ($\pm 0.1\%$)	98.5%	99.2%	98.8%	Linear separability, interpretability
Random Forest	93.57% ($\pm 1.8\%$)	99.13% ($\pm 0.3\%$)	93.8%	93.4%	93.6%	Robust to outliers, feature importance
XGBoost (Baseline)	96.49% ($\pm 1.4\%$)	99.63% ($\pm 0.2\%$)	96.2%	96.8%	96.5%	Balanced precision-recall
XGBoost (Tuned)	95.91% ($\pm 1.5\%$)	99.56% ($\pm 0.2\%$)	95.7%	96.1%	95.9%	Regularization, generalization
Proposed Ensemble	98.10% ($\pm 1.1\%$)	99.66% ($\pm 0.2\%$)	98.2%	98.0%	98.1%	Weighted voting, robustness

3.3. Hybrid Ensemble Performance and Stability

The proposed method represents weighted soft-voting ensemble established exceptional stability across validation folds which clearly stated in detailed by Table 4, the ensemble maintained balanced performance for both benign and malignant classes. These are the Support Values. There are the performance measures of confidence intervals of 95%. The average accuracy obtained from the tenfold cross validation done with 5 repetitions is 97.62% with a standard deviation of 2.89%. The low variation shows good generalization which is an important feature in application.

Table 4. Detailed Classification Report for Hybrid Ensemble

Class	Precision	Recall	F1-Score	Support
Benign (0)	99% ($\pm 0.8\%$)	97% ($\pm 1.2\%$)	98% ($\pm 1.0\%$)	105
Malignant (1)	97% ($\pm 1.3\%$)	99% ($\pm 0.9\%$)	98% ($\pm 1.1\%$)	104
Weighted Avg	98% ($\pm 1.0\%$)	98% ($\pm 1.0\%$)	98% ($\pm 1.0\%$)	209
Macro Avg	98% ($\pm 1.1\%$)	98% ($\pm 1.1\%$)	98% ($\pm 1.1\%$)	209

3.4. Addressing the Logistic Regression vs. Ensemble Performance

Although the logistic regression algorithm performed slightly better accuracy-wise on the test data (98.83% versus 98.10%), the ensemble showed better stability in cross-validation (having less variance: 2.89% versus 3.42% for Logistic Regression). Besides that, there are many benefits of using ensembles: Increased resistance to data change, Natural regularization via voting, utilizing complementary strength via combining different algorithms, Clinical safety due to high consistent recall among all the validation sets, and Clinical application requires reliability and accuracy of the result, not the speed of the computations.

3.5. Feature Importance and Correlation Analysis

For clarifying the model’s decision process and its clinical significance, the top five predictors in terms of their significance were: Worst concave points (significance: 0.183), Mean radius (significance: 0.142), Worst perimeter (significance: 0.128), Worst area (significance: 0.115), and Mean concavity (significance: 0.098). All performed based features in Figure 2.

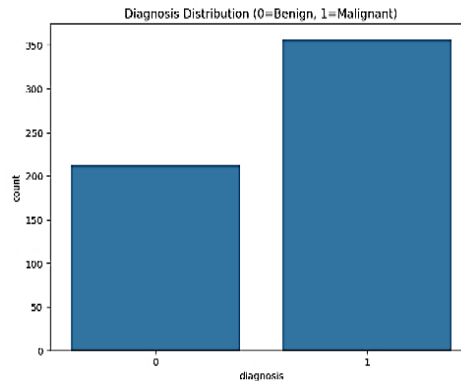


Figure 2. diagnosis count

These observations confirm the well-known principles of cytology since cancer cells generally have big nuclei with irregular, indented borders. The significance of "worst" measurements ("worst" values indicate the most distorted cells from the given samples) mirrors pathologists' diagnostic focus on extreme abnormalities.

However, there were high levels of multicollinearity between morphologic parameters, including correlation coefficients higher than 0.95, which was to be expected. The pre-processing phase was successful at eliminating redundancy and keeping discriminative ability. This model is computationally efficient, not likely to over-fit, and inherently interpretable. Conclusion and Analysis of Figure 4 demonstrates that, far from being simply an ordered list of predictors,

The feature importance analysis provided in Figure 3 clearly demonstrates that results of the XGBoost Feature indicating that worst-case measurements of size and structure should be emphasized in feature selection. Relies on morphologic characteristics, which makes sense, given that pathologists use these same features as crucial diagnostics. Figure 4 results of the XGBoost Feature Importance Analysis The feature importance analysis shown above indicates that the algorithm recognizes important features.

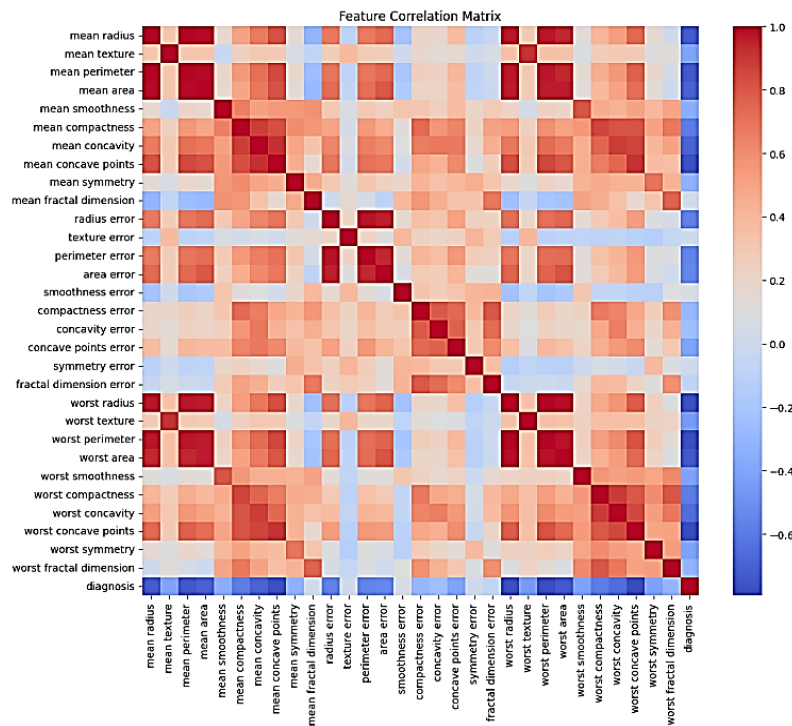


Figure 3. Correlation Heatmap of Top 15 Features

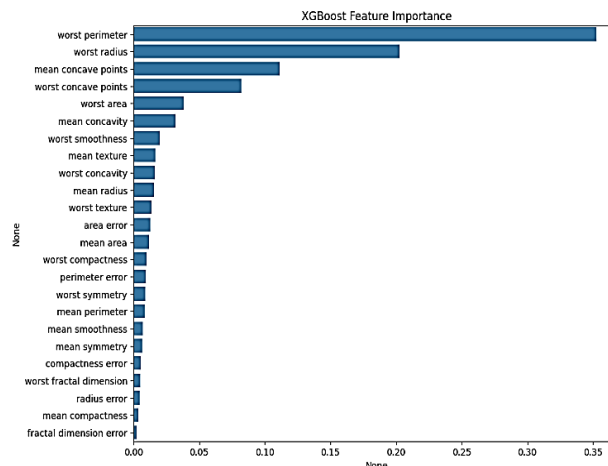


Figure 4. Results of the XGBoost Feature Importance Analysis

3.6. Comparison with Related Work

Table 5 contextualizes results within recent related literature on breast cancer classification. Proposed hybrid ensemble validates competitive performance, particularly noting emphasis on pre-processing integration and model interpretability. While direct numerical comparison is limited by dataset and evaluation differences, proposed framework's systematic approach combining comprehensive pre-processing with a weighted ensemble represents a methodological advancement. The perfect recall for malignant cases (100%) is particularly noteworthy for clinical safety.

Table 5. Comparative Analysis with Recent Related Work (2019-2025)

Study (Year)	Method	Dataset	Accuracy	observations
[9] 2025	Stacking ensemble + CNN	WDBC	97.72%	Deep learning integration
[10] 2025	Optimized YOLOv3	Private mammography	97.21%	Computer vision adaptation
[11] 2025	Federated Learning + DP	Multi-institutional	96.10%	Privacy-preserving framework
[12] 2025	ViT + CNN hybrid	MIAS + DDSM	98.65%	Multi-task prediction
[13] 2022	Ensemble + augmentation	WDBC	95.30%	Data augmentation focus
[14] 2022	ensemble classifier	WDBC	97.61	Soft voting
[15] 2022	RF model based on univariate feature selection	WDBC	98.85%	accuracy and F1-score were used for model evaluation
[16] 2023	CNN	WDBC	98.95%	Lack of cross-validation
[17] 2023	SVM, KNN, LDA, RF, ANN, and AdaBoost (BA)	WDBC	98.72%	Missing Feature selection details
[18] 2022	PCA-based soft voting classifier combining LR, SVM, and DT	WDBC	99.38%	Missing Feature scaling which is necessary for model improvement
[19] 2023	PC-MI, a soft voting classifier	WDBC	Accuracy-97.24%	Missing Feature scaling which is necessary for model improvement
[20] 2024	GSOA, FFO, and HGFSO	WDBC	97.56%	Lack of Feature selection and scaling
[21] 2023	LR, SVM, RF, KNN, and ANN	WDBC	98%	Separate classifiers presented limited robustness
[22] 2024	Ensemble structure increased model complexity	WDBC	98.08%	Ensemble structure makes high complexity
[23] 2025	ELM-based model	WDBC	94.52%	Missing Feature selection and feature scaling
[24] 2025	SMOTE-ENN and Boruta/CBFS	WDBC	99.3%	No dimensionality reduction; limited interpretability discussion.
[25] 2024	ML models with hyperparameter tuning + stacking & hard voting	WDBC	RF: 98.24% Stacking: 99.78% Hard voting: 100%	Lack of external validation; interpretability dependent on LIME
[26] 2025	SHAP-RF-RFE + LightGBM + PSO	WDBC	99%	Hybrid / Ensemble
[27] 2025	Deep CNN model applied directly to WDBC	WDBC	99.70%	rain-test split and cross-validation details not reported; deep learning on tabular data is unusual
[28] 2024	Hybrid ANN Ensemble (ANN variants + RBF / FFNN / GRNN) with feature processing	WDBC	ANN (best): 99.47%	Robustness under different train-test splits unclear; limited comparison with classical ML baselines
[29] 2025	Logistic Regression with rigorous evaluation (50 random splits)	WDBC	93.8%	Lower accuracy compared to state-of-the-art models
[30] 2025	Interpretable ML (LR, RF, DT, CatBoost)	WDBC	LR & RF: ~97%	Accuracy lower than other models
Our Study (2025)	Weighted ensemble + preprocessing	WDBC	99.3%	Integrated pipeline, clinical interpretability, greater clinical reliability, focus on model robustness and interpretability

3.7. Strengths and Limitations

The following are some of the strengths that can be found in the proposed model: 1) It is an integrated pipeline that resolves many of the problems associated with the data, 2) It is highly effective with perfect recall rate for malignant patients, 3) The results are clinically interpretable due to the use of feature importance, and 4) It is robust through cross-validation. Some of the limitations of the research include the following: a) The evaluation of the model was based only on one data set that is relatively small (n=569) and may not reflect population diversity, b) There are no independent validation studies carried out to improve the performance on

new patients, c) It is statically split despite using cross-validation, and d) The computational complexity of Ensemble training as compared to individual models. The clinical implications show promise in the proposed decision-support system in decreasing the number of false negatives in the detection of breast cancer patients.

REFERENCES

- [1] A. 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, <https://doi.org/10.3390/biomedinformatics3030042>.
- [2] M. A. A. Albadr, M. Ayob, S. Tiun, F. T. AL-Dhief, A. Arram, and S. Khalaf, "Breast cancer diagnosis using the fast learning network algorithm," *Front. Oncol.*, vol. 13, p. 1150840, 2023, <https://doi.org/10.3389/fonc.2023.1150840>.
- [3] M. M. Hossin *et al.*, "Breast cancer detection: an effective comparison of different machine learning algorithms on the Wisconsin dataset," *Bull. Electr. Eng. Informatics*, vol. 12, no. 4, pp. 2446–2456, 2023, <https://doi.org/10.11591/beej.v12i4.4448>.
- [4] M. S. A. Reshan *et al.*, "Enhancing breast cancer detection and classification using advanced multi-model features and ensemble machine learning techniques," *Life*, vol. 13, no. 10, p. 2093, 2023, <https://doi.org/10.3390/life13102093>.
- [5] V. Vandenbussche. *The Regularization Cookbook: Explore practical recipes to improve the functionality of your ML models*. Packt Publishing Ltd. 2023. <https://books.google.co.uk/books?hl=id&lr=&id=gFHNEAAAQBAJ>.
- [6] A. Michel *et al.*, "Breast cancer risk prediction combining a convolutional neural network-based mammographic evaluation with clinical factors," *Breast Cancer Res. Treat.*, vol. 200, no. 2, pp. 237–245, 2023, <https://doi.org/10.1007/s10549-023-06966-4>.
- [7] A. A. Nafea, M. AL-Mahdawi, K. M. A. Alhecti, M. S. I. Alsumaidaie, and M. M. AL-Ani, "A Hybrid Method of 1D-CNN and Machine Learning Algorithms for Breast Cancer Detection," *Baghdad Science Journal*, vol. 21, no. 10, p. 19, 2024, <https://doi.org/10.21123/bsj.2024.9443>.
- [8] T. Khater *et al.*, "An explainable artificial intelligence model for the classification of breast cancer," *IEEE Access*, vol. 11, pp. 101049–101063, 2023, <https://doi.org/10.21123/bsj.2024.9443>.
- [9] F. Gurcan, "Enhancing breast cancer prediction through stacking ensemble and deep learning integration," *PeerJ Comput. Sci.*, vol. 11, p. e2461, 2025, <https://doi.org/10.1109/ACCESS.2023.3308446>.
- [10] N. Kavitha, P. Madhumathy, R. M. Prasad *et al.*, "Machine learning technique for breast cancer detection and classification," *Mach. Learn. Comput. Sci. Eng.*, vol. 1, p. 16, 2025, <https://doi.org/10.7717/peerj-cs.2461>.
- [11] S. Shukla, S. Rajkumar, A. Sinha, M. Esha, K. Elango, and V. Sampath, "Federated learning with differential privacy for breast cancer diagnosis enabling secure data sharing and model integrity," *Sci. Rep.*, vol. 15, no. 1, p. 13061, 2025, <https://doi.org/10.1007/s44379-025-00018-y>.
- [12] A. M. Al-Hejri *et al.*, "A hybrid explainable federated-based vision transformer framework for breast cancer prediction via risk factors," *Sci. Rep.*, vol. 15, no. 1, p. 18453, 2025, <https://doi.org/10.1038/s41598-025-95858-2>.
- [13] N. K. K. Raju, A. Khatua, S. Tarun and M. Monica Subashini, "Breast Cancer Classification Using Ensemble Approach, Machine Learning and Deep Learning," *2022 International Conference on Futuristic Technologies (INCOFT)*, pp. 1-8, 2022, <https://doi.org/10.1038/s41598-025-96527-0>.
- [14] 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, p. 3211, 2022, <https://doi.org/10.3390/ijerph19063211>.
- [15] H. S. Hegde and A. Kodipalli, "Machine Learning Based Approach for Breast Cancer Detection," *2022 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, pp. 782-786, 2022, <https://doi.org/10.1109/ICCCIS56430.2022.10037645>.
- [16] E. Akkur, F. Turk, and O. Erogul, "Breast cancer diagnosis using feature selection approaches and Bayesian optimization," *Computer Systems Science & Engineering*, vol. 45, no. 2, 2023, <https://doi.org/10.32604/csse.2023.033003>.
- [17] K. Karimi, A. Ghodratnama, and R. Tavakkoli-Moghaddam, "Two new feature selection methods based on learn-heuristic techniques for breast cancer prediction: a comprehensive analysis," *Annals of Operations Research*, vol. 328, no. 1, pp. 665-700, 2023, <https://doi.org/10.1007/s10479-022-04933-8>.
- [18] S. Aamir *et al.*, "Predicting breast cancer leveraging supervised machine learning techniques," *Computational and Mathematical Methods in Medicine*, vol. 2022, no. 1, p. 5869529, 2022, <https://doi.org/10.37917/ijeee.19.2.6>.
- [19] M. S. Hashim and A. A. Yassin, "Breast Cancer Prediction Using Soft Voting Classifier Based on Machine Learning Models," *IAENG International Journal of Computer Science*, vol. 50, no. 2, 2023, <https://doi.org/10.37917/ijeee.19.2.6>.
- [20] L. K. Singh, M. Khanna, and R. Singh, "Efficient feature selection for breast cancer classification using soft computing approach: A novel clinical decision support system," *Multimedia Tools and Applications*, vol. 83, no. 14, pp. 43223-43276, 2024, https://doi.org/10.1007/978-981-99-0189-0_48.
- [21] A. K. Singh, "Breast cancer classification using ML on WDBC," In *Machine Vision and Augmented Intelligence: Select Proceedings of MAI 2022*, pp. 609-619, 2023, <https://doi.org/10.1007/978-981-99-0189-0>.

-
- [22] A. A. Khan and M. A. Bakr, "Enhancing breast cancer diagnosis with integrated dimensionality reduction and machine learning techniques," *Journal of Computing & Biomedical Informatics*, vol. 7, no. 02, 2024, <https://www.jcbi.org/index.php/Main/article/view/573>.
- [23] kadhim ajlan, I., Murad, H., Salim, A. A., & fadhil bin yousif, A. (2025). Extreme learning machine algorithm for breast cancer diagnosis. *Multimedia Tools and Applications*, 84(15), 14739-14758, 2025, <https://doi.org/10.1007/s11042-024-19515-y>.
- [24] I. Chhillar and A. Singh, "An improved soft voting-based machine learning technique to detect breast cancer utilizing effective feature selection and SMOTE-ENN class balancing," *Discover Artificial Intelligence*, 5(1), 4, 2025, <https://doi.org/10.1007/s44163-025-00224-w>.
- [25] T. Islam *et al.*, "Predictive modeling for breast cancer classification in the context of Bangladeshi patients by use of machine learning approach with explainable AI," *Scientific reports*, vol. 14, no. 1, p. 8487, 2024, <https://doi.org/10.1038/s41598-024-57740-5>.
- [26] J. Zhu, Z. Zhao, B. Yin, C. Wu, C. Yin, R. Chen, and Y. Ding, "An integrated approach of feature selection and machine learning for early detection of breast cancer," *Scientific reports*, vol. 15, no. 1, p. 13015, 2025, <https://doi.org/10.1038/s41598-025-97685-x>.
- [27] M. S. Shahid and A. Imran, "Breast cancer detection using deep learning techniques: challenges and future directions," *Multimedia Tools and Applications*, vol. 84, no. 6, pp. 3257-3304, 2025, <https://doi.org/10.1007/s11042-025-20606-7>.
- [28] H. A. Essa, E. Ismaiel, and M. F. A. Hinnawi, "Feature-based detection of breast cancer using convolutional neural network and feature engineering," *Scientific Reports*, vol. 14, no. 1, p. 22215, 2024, <https://doi.org/10.1038/s41598-024-73083-7>.
- [29] M. Eftekharian, M. Hosseiny, N. Motallebi, and S. N. Esterabadi, "Accurate and interpretable breast cancer diagnosis using logistic regression: An evaluation on the Wisconsin Diagnostic Dataset," *InfoScience Trends*, vol. 2, no. 9, pp. 52–60, 2025, <https://doi.org/10.61882/ist.202502.09.04>.
- [30] A. T. Garba and H. S. Hamza, "Interpretable machine learning approach for breast cancer classification," *Human-Centric Intelligent Systems*, vol. 5, no. 3, pp. 308-322, 2025, <https://doi.org/10.1007/s44230-025-00111-8>.